BOOK 2

SMstudy
Targeting success

SMstudy® Guide

MARKETING RESEARCH

2017 Edition

A Guide to the SMstudy® Sales and Marketing Body of Knowledge (SMstudy® Guide)

SMstudy® Guide

MARKETING RESEARCH

2017 Edition

A Comprehensive Guide to Marketing Research

ISBN: 978-0-9899252-6-6

Library of Congress Control Number: 2016920774

Published by:
SMstudy®, a brand of VMEdu®, Inc.
12725 W. Indian School Road,
Suite F-112, Avondale,
AZ – 85392, USA
Phone: +1-480-882-0706
Fax: +1-240-238-2987
E-mail: smbok@smstudy.com
Website: www.smstudy.com

SMstudy® welcomes corrections and comments on its books. Please feel free to send comments on typographical, formatting, or other errors. You can make a copy of the relevant page of the book, mark the error, and send it to the above address or send an e-mail to smbok@smstudy.com.

10 9 8 7 6 5 4 3 2

PREFACE

The SMstudy® Guide (also referred to as a Guide to the Sales and Marketing Body of Knowledge, or *SMBOK® Guide)* is a comprehensive process-oriented framework for the planning and execution of activities associated with all facets of Sales and Marketing. It can be used as a reference source for experienced Sales and Marketing professionals or as a detailed guide for individuals or students with little prior Sales and Marketing knowledge or experience. The *SMstudy® Guide* can be applied in any organization or industry—from small companies with only a few employees to large, complex organizations with numerous business units, multiple product lines, and thousands of employees across many countries.

The *SMstudy® Guide* series consists of six books for all the aspects of Sales and Marketing: Marketing Strategy, Marketing Research, Digital Marketing, Corporate Sales, Branding and Advertising, and Retail Marketing. Each book addresses a key component of Marketing and can be used as a standalone resource or as part of a more comprehensive program that utilizes any number of the six books as required by the business. In this way the *SMstudy® Guide* series offers a flexible framework that can be tailored to address the specific needs of each organization

This third book, Marketing Research, provides a framework for conducting Marketing Research for a company's products, services, and brands. Marketing Research is the systematic process of collecting, processing, and analyzing data to provide required information to decision makers. Marketing Research is linked to all other Aspects of Marketing as it provides critical insights that inform key decisions in all other marketing planning and strategies. The processes associated with planning and executing a marketing research project include understanding the research problem; deciding a suitable research design; collecting, processing, and analyzing data required to solve the problem; interpreting data; and presenting the findings/recommendations of the research project to the key decision makers.

I would like to thank the 42 authors, subject matter experts, and reviewers who greatly contributed to the creation of this body of knowledge. Their combined efforts and collaborations have resulted in a comprehensive, highly effective, and unique approach to understanding, planning, and implementing Sales and Marketing initiatives.

Tridibesh Satpathy
CEO, SMstudy®

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

A Guide to the SMstudy® Sales and Marketing Body of Knowledge (SMBOK® Guide), also referred to as the "*SMstudy® Guide,*" is a series of books that provide guidelines for the Sales and Marketing of products and services. It offers a comprehensive framework that can be used to effectively manage Sales and Marketing efforts in any organization. The objective of the *SMstudy® Guide* is to provide a practical and process-oriented approach to Sales and Marketing that emphasizes how various elements of Sales and Marketing can be integrated to develop a comprehensive and effective organizational Sales and Marketing Plan.

The concepts in the *SMstudy® Guide* can be applied effectively to any company in any industry—from small companies with only a few employees to large, complex organizations with numerous business units, multiple product lines, and thousands of employees across many countries. The term "product" in the *SMstudy® Guide* may refer to either a product or a service of a company.

This introductory chapter includes definitions of key terms, the purpose and framework of the *SMstudy® Guide*, an overview of the Aspects of Sales and Marketing discussed throughout the *SMstudy® Guide*, the SMstudy® certification schema, a brief history of Marketing Research, and a general overview of the contents of this book on Marketing Research. This chapter also briefly discusses Corporate Strategy and its relationship to Sales and Marketing.

This chapter is divided into the following sections:

1.1 How to Use the *SMstudy® Guide?*

1.2 Why Use the *SMstudy® Guide*?

1.3 A Brief History of Marketing Research

1.4 Corporate Strategy Overview

1.5 Aspects of Sales and Marketing

1.6 The Levels of Sales and Marketing Strategy

1.7 Marketing Strategy Overview

1.8 Marketing Research Overview

1.1 How to Use the SMstudy® Guide?

The *SMstudy® Guide* can be used as a reference and knowledge guide by experienced Sales and Marketing practitioners, as well as by persons with little prior knowledge or experience in Sales or Marketing roles. Because the *SMstudy® Guide* offers a comprehensive Sales and Marketing framework, many will find value in using this resource to guide decision-making and planning across all facets of Sales and Marketing; however, the contents of the Guide are organized to enable quick and easy reference for individuals who may be interested in, or studying, only one or two specific facets of Sales or Marketing. Similarly, the *SMstudy® Guide* provides a valuable tool for individuals already in distinct Sales or Marketing roles (e.g., Digital Marketing Manager), as its design enables such individuals to focus on the specific Aspects that are most relevant to such roles.

1.1.1 Process-Oriented Approach with Defined Inputs, Tools, and Outputs

In order to facilitate the best application of the *SMstudy® Guide* framework, the *SMstudy® Guide* defines a process-oriented approach to Sales and Marketing, which provides specific guidance to Sales and Marketing professionals about how to most effectively and efficiently manage their marketing activities. The *SMstudy® Guide* defines Sales and Marketing in terms of processes that comprise a series of actions that lead to a particular result. Each process requires specific inputs and then uses various tools to create specific outputs. To cater to the needs of a diverse audience with varying levels of expertise in Sales and Marketing, the *SMstudy® Guide* has differentiated highly recommended inputs, tools, and outputs from recommended, but optional ones. Inputs, tools, and outputs denoted by asterisks (*) are highly recommended, while others with no asterisks are recommended, but optional. It is suggested that those individuals being introduced to Sales and Marketing focus primarily on the highly recommended inputs, tools, and outputs for each process, while more experienced practitioners should thoroughly understand all of the relevant inputs, tools, and outputs for each process.

1.1.2 Using SMstudy® Guide with SMstudy.com Website and VMEdu® Mobile App

The SMstudy.com website and the VMEdu® mobile app provide additional resources to help individuals better understand and apply the Sales and Marketing framework defined in the *SMstudy® Guide.* The website and app include the following:

- A certification schema, which helps students study marketing subjects in a structured manner, get tested on relevant concepts through proctored certification exams, and gain relevant certifications which demonstrate their knowledge and experience in different areas of Sales and Marketing (see section 1.1.3 for a description of the certification schema)
- High-quality videos with relevant and interesting examples that help individuals gain a thorough understanding of specific concepts

- Case studies that illustrate how the *SMstudy® Guide* framework can be used in real-life scenarios
- Additional resources for students to obtain expert training through physical classrooms, virtual instructor-led sessions, and high-quality online courses
- A glossary of terms, flashcards, study guides, and more

1.1.3 Certification Schema for SMstudy® Certifications

The certifications related to the *SMstudy® Guide* are managed by SMstudy.com. The certification schema is outlined in Figure 1-1.

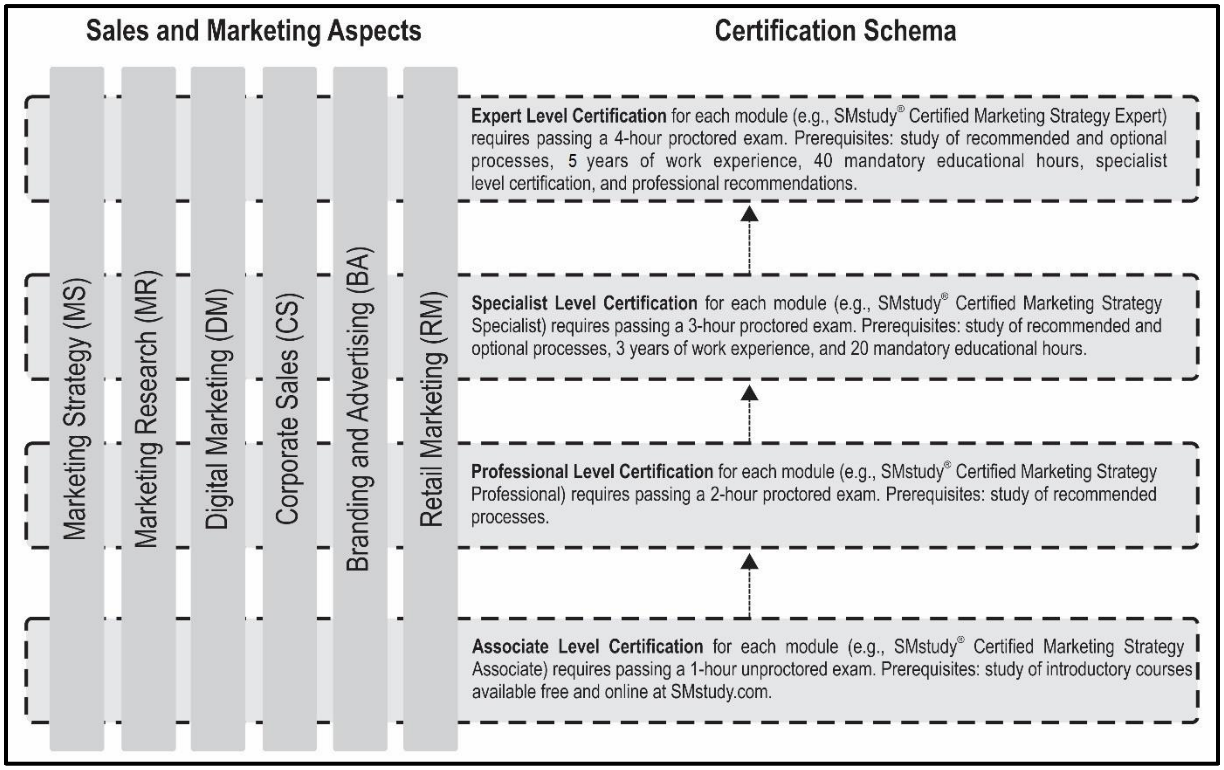


Figure 1-1: SMstudy® Certification Schema

The following is a brief description of each level of certification:

- **Associate Level Certifications**—The introduction modules are available at no charge to interested individuals. All Aspects of Sales and Marketing have an applicable Associate-level certification (e.g., “SMstudy® Certified Marketing Strategy Associate”). The certification exams are free and not proctored; and candidates have one hour to complete each exam. The prerequisite is an understanding of the highly recommended inputs, tools, and outputs for each process relevant to the particular Aspect of Sales and Marketing. There is no work experience requirement and no mandatory educational hours in addition to the recommended study.

- **Professional Level Certifications**— All Aspects of Sales and Marketing have an applicable Professional-level certification (e.g., "SMstudy® Certified Marketing Strategy Professional").The certification exams are proctored and candidates have two hours to complete each exam. The prerequisite is a study of the relevant *SMstudy® Guide* book with more emphasis on the highly recommended inputs, tools, and outputs for each process relevant to the particular Marketing Aspect. There is no work experience requirement and no mandatory educational hours in addition to the recommended study. Individuals who pass the certification exams for three or more Professional modules are awarded an additional certification called "SMstudy® Certified Sales and Marketing Professional."

- **Specialist Level Certifications**—All Aspects of Sales and Marketing have an associated Specialist-level certification (e.g., "SMstudy® Certified Marketing Strategy Specialist"). The certification exams are proctored and candidates have three hours to complete each exam. The prerequisites are a study of all of the relevant inputs, tools, and outputs for each process, three years of related work experience, and twenty mandatory educational hours. Individuals who pass the certification exams for three or more Specialist modules are awarded an additional certification called "SMstudy® Certified Sales and Marketing Specialist."

- **Expert Level Certifications**—All Aspects of Sales and Marketing have an associated Expert-level certification (e.g., "SMstudy® Certified Marketing Strategy Expert"). The certification exams are proctored and candidates have four hours to complete each exam. The prerequisites are attaining the Specialist level certification for that specific Aspect, a study of all of the relevant inputs, tools, and outputs for each process, five years of related work experience, forty mandatory educational hours, and recommendations from two peers and a manager. Individuals who pass the certification exams for three or more Expert modules are awarded an additional certification called "SMstudy® Certified Sales and Marketing Expert."

Other than the certifications mentioned above, SMstudy® offers additional certifications in fields related to Sales and Marketing such as Affiliate Marketing, E-mail Marketing, Search Engine Optimization, Search Marketing, Social Media and Web Analytics. Information about these certifications is available in the SMstudy.com website.

1.2 Why Use the SMstudy® Guide?

Some of the key benefits of using the *SMstudy® Guide* are as follows:

1. **Consolidated Expertise**—The *SMstudy® Guide* is developed by VMEdu, Inc., a global certification course provider that has educated over 400,000 students world-wide in more than 3,500 companies. It provides practical, industry-proven best practices, rather than purely theoretical advice.

2. **Process-Oriented Approach**—The *SMstudy® Guide* explains Sales and Marketing concepts through a practical, process-oriented approach. This helps Sales and Marketing professionals understand the specific processes they should follow to be effective in their Sales and Marketing roles. Each process has associated inputs, tools, and outputs that are recommended for use. Highly recommended inputs, tools, and outputs are noted with an asterisk (*) beside the concept in each process box and then again when each process is discussed throughout that section.

3. **Applicable to All Industries**—The many authors, advisers, and reviewers of the *SMstudy® Guide* have worked in numerous Sales and Marketing areas and geographic regions across a variety of industries. Thus, insights provided by them make this body of knowledge industry independent.

4. **Applicable to Companies of All Sizes**—The *SMstudy® Guide* has been written to meet the needs of all companies regardless of size. Small start-up companies with fewer than ten employees, or large organizations with several thousand employees and multiple product lines and business units, can equally benefit from the information in this guide. Additionally, the content provided in the *SMstudy® Guide* is highly relevant to for-profit and non-profit organizations alike.

5. **Comprehensiveness**—The *SMstudy® Guide* is organized into six Sales and Marketing Aspects: Marketing Strategy (MS), Marketing Research (MR), Digital Marketing (DM), Corporate Sales (CS), Branding and Advertising (BA), and Retail Marketing (RM). Each Aspect is detailed in a separate book. Taken together, the series provides a comprehensive and complete understanding of Sales and Marketing. The concepts covered in the *SMstudy® Guide* are further reinforced through videos and case studies available at SMstudy.com.

6. **Applicable to Beginners and Experts**—The *SMstudy® Guide* presents recommended concepts that beginners should know and also highlights advanced concepts for individuals who have more experience and who are on their way to becoming experts in the field. Readers can decide which of the six Sales and Marketing Aspects are most relevant to them and select from the available books accordingly.

7. **Alignment with Job Roles**—The Aspects included in the *SMstudy® Guide* are organized to align with the most common or typical job roles or career fields of Sales and Marketing.

8. **Continuous Improvement**—Concepts related to Sales and Marketing continue to evolve; therefore, the *SMstudy® Guide* will be continuously reviewed and updated to ensure that it remains relevant.

1.3 A Brief History of Marketing Research

Although Sales and Marketing evolved many centuries ago with the advent of the barter system, Marketing Research as a practice developed relatively recently. The use of Marketing Research in Sales and Marketing can be traced to the evolution of conventional mass media marketing. Unlike a seller's marketplace where sellers had the advantage over customers, mass media marketing featured multiple manufacturers, thus shifting the balance of power in favor of consumers. Manufacturers created differentiated perceptions for their products by developing brands or names for their specific products or services with a specific message or positioning. They also began advertising their products or brands for a wider reach. With the introduction of brands and advertising, manufacturers felt the need to understand how their brands and advertisements performed among consumers. Early Marketing Research started in the form of surveys and interviews that gauged people's recall of ads that appeared in print media. The scope of Marketing Research increased with the progress in conventional mass media marketing and later with the advent of advanced models of Sales and Marketing. The following section discusses the evolution of Sales and Marketing models over time and describes the role that Marketing Research played in each of them.

1.3.1 Early Models of Sales and Marketing

Sales and Marketing has evolved significantly over time, starting over one thousand years ago with the simple barter system and spanning several centuries, adapting to changes in consumer behavior, modes of communication, and advancements in technology, in order to become the multifaceted, multimedia discipline that it is today. It is important to take a look at the interesting history of Sales and Marketing in order to gain an understanding and appreciation for the systems and concepts that are used today to promote goods and services to consumers.

Figure 1-2 depicts the timeline for the Evolution of Sales and Marketing; a comprehensive description of the history of Sales and Marketing can be found in the *SMstudy® Guide-Book 1, Marketing Strategy* (section 1.3)

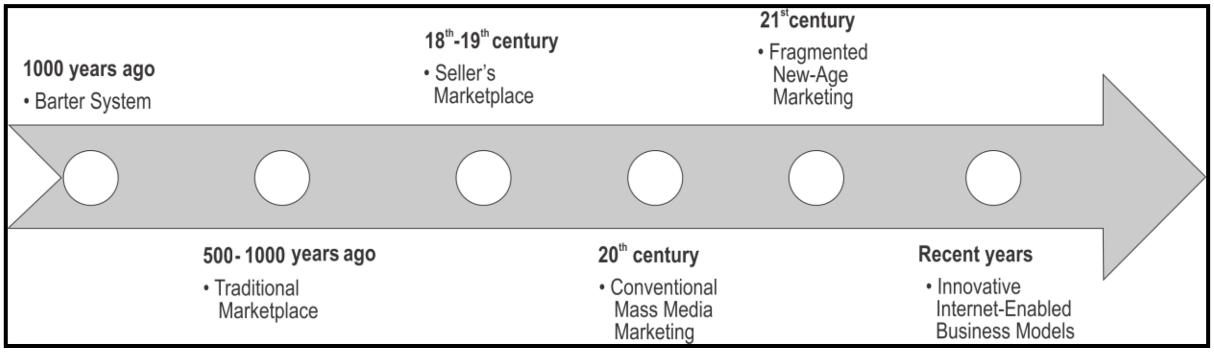


Figure 1-2: Evolution of Sales and Marketing Timeline

Marketing began several centuries ago with the barter system whereby various goods and services were exchanged for other goods and services. This concept of exchange is the foundation of Sales and Marketing and, while the barter system is less common today, the dawn of the Internet in the 1990s made this simple form of transaction even more convenient by offering a means for non-local individuals to match their needs with the offerings of others, both locally and across the globe. While the barter system introduced consumers to the concept of exchange, the introduction of various forms of currency several centuries later, made the concept of exchange a simpler proposition. No longer was a match of ‘needs’ required—goods and services could be exchanged for money—and the “Traditional Marketplace” was born. Over the years, the concept of mass production as well as supply chain and distribution channels were introduced giving way to the “Seller’s Marketplace.” During this time, the transportation infrastructure and communication systems substantially improved, which allowed factories to mass produce and sell the products to a wide variety of customers, a concept often referred to as “mass marketing.” Although not directly applied in Sales and Marketing, various research techniques that would later be used for Marketing Research evolved during this period. The early predecessors of Marketing Research included surveys and studies conducted by governments and agencies for analyzing economic, social, and political factors that affected people. Censuses, poll surveys and studies for production thus were precursors to modern Marketing Research techniques.

1.3.2 Conventional Mass Media Marketing

The number of manufacturers or industries grew in the twentieth century giving consumers more options to buy from multiple manufacturers. Manufacturers started creating differentiated perceptions for their products which led to the concept of branding. Various channels such as print advertising (newspapers, magazines, inserts, or run of paper) mass mailers (flyers, postcards), television (network, cable, or syndication), radio (national, local, satellite, or podcast), and outdoor advertising emerged through which products were advertised and promoted. These developments shifted the balance of power in favor of consumers. The objective of conventional mass media marketing is for organizations to create strong brands and differentiated brand perceptions so that consumers will desire and purchase their products rather than those available from competitors. Thus, mass media marketing usually uses cumulative repetition over time to influence consumer attitudes and purchase actions. Mass media marketing also involves creating distribution channels and appropriate pricing and positioning strategies to ensure that desirable products are available to customers at specific price points.

Marketing Research evolved as an important aspect for mass media marketing. Manufacturers felt the need to identify customer preferences across regions and demographics and create or customize products that suited a particular target segment. Early Marketing Research relied primarily on surveys and interviews through which customer perceptions, preferences, attitudes, lifestyles etc. were gauged. The information gathered was used to make decisions regarding product attributes and branding. The advances in mass media allowed industries to increase the reach and popularity of their brands as well as increase the scope of conducting market research.

1.3.3 Fragmented New-Age Marketing

In recent times, the media has become increasingly fragmented with several hundred television and radio channels, as well as a large variety of print media, including newspapers, magazines, and trade publications. Moreover, since the late nineteen nineties, with the increasing popularity of the Internet and, more recently, smartphones, many options now exist for advertisers to reach a global audience using digital media marketing methods. With all of these options, many marketers find it beneficial to use an integrated approach to marketing by leveraging the strengths of various types of media. Companies must evaluate all media in terms of who the target audience is and what media resonates with them best. In many cases, assumptions will need to be made and incorporated into the media-testing framework (e.g., for each planning period a company might allocate a certain amount of its marketing budget to test new methods).

Some characteristics of fragmented new-age marketing are as follows:

- It is a fact that people now spend more time on the Internet using smartphones, tablets, or computers than they spend through conventional mass media, such as television, radio, or newspapers. This is especially true for the thirty-year-old and younger market segment. Since Sales and Marketing is most successful when it meets the demands of consumers, this change in consumer preferences is significantly altering the Sales and Marketing landscape for established companies. Businesses are discovering that conventional mass media marketing has limited effectiveness and some customer segments are not even reachable using these traditional media forms.

- Fragmented new-age marketing generally supports new, small brands with much smaller budgets targeted directly to customers in a global marketplace. This represents a significant distinction from conventional mass media marketing, where achieving a global reach for a small company may have been prohibitively expensive.

- While mass media marketing is less targeted and primarily focused on affecting emotional attitudes about the brand, new-age marketing is data-driven and more focused on driving specific calls to action. Also, while mass media marketing typically involves interruption (e.g., people watching a television program which is "interrupted" by an advertisement), new-age marketing is about engagement (e.g., offering relevant content that is of value to people).

- Unlike older media options where Sales and Marketing communications were primarily uni-directional (i.e., from producers to end-consumers), communications have increasingly become multi-directional (i.e., from producers to consumers, consumers to producers, and consumers to consumers). For example, there are multiple rating websites available where customers can provide independent ratings of a company’s products or services, and others, including the company itself, can respond or elaborate on these ratings. Although generally a benefit to both producers and consumers, this trend can make brand management challenging for companies if actual or potential customers perceive that a product does not reflect the brand message intended by marketing efforts.

- Due to the nature of new-age marketing, consisting of multiple media forms and the ability to generate significant information, huge amounts of data (commonly referred to as “big data”) are now available to companies. The ability to process this data through proper marketing analytics, and assimilate such data to generate valuable insights, can become a significant differentiator for ensuring that companies engage in “smart marketing” (i.e., to generate greater revenues with relatively smaller marketing budgets).

Marketing Research evolved significantly during this period with the advances in computing power and statistical techniques, giving rise to tools that allowed large amounts of data to be analyzed and interpreted with relative ease. The use of computers reduced the effort required for collecting and analyzing data, at the same time increasing accuracy and reducing costs. During this period Marketing Research graduated from being just a supporting activity for decision makers. Market researchers were now expected to provide intelligent insights that were interpreted by analyzing data. Companies now relied on Marketing Research to decide on key strategies for all Aspects of Sales and Marketing. The advent of the Internet provided more options for companies to reach out to the target audience as well as for the public to interact, discuss, and provide feedback about the company’s products. The use of Internet-based multi-directional channels such as social media also made it possible for companies to collect feedback for marketing campaigns run across any channel.

Examples of Fragmented New-Age Marketing:

- Social media (e.g., Facebook, Google+, YouTube) and company websites allow small companies to showcase their products at a low cost or at times, even for free. Companies can share engaging content, which can go viral, thus promoting their brand and reaching a global audience. Brands can also produce informative, instructional content via blog posts, forums, and so forth.
- Online paid advertising (e.g., Google AdWords, LinkedIn Sponsored Updates, Facebook Ads) allows companies to market their products or services to their target audiences at smaller budgets compared to conventional mass media marketing. For many companies, online paid advertising is replacing conventional mass media marketing. It may be important to note that unlike mass media, which is a one-way broadcast from brand to consumer, new-age marketing involves a two-way interaction between brand and consumer. For example, in the case of Google AdWords, the consumer clicks on an ad that takes him to the landing page of the brand's website. At that point, the brand will likely ask that the consumer do something (i.e., a specific call to action to download material or place an order). If the consumer accepts the call to action, the company provides additional information on the product itself. Thus, the transaction is usually a back-and-forth engagement between the company and its potential customers.
- Multi-directional communication is facilitated using fragmented new-age marketing. Customers can provide their own feedback or research what others have to say about particular products through blogs, Twitter feeds, Google reviews, and so on. TripAdvisor provides a forum for travel enthusiasts to share feedback about their experiences staying in particular hotels. Such feedback can significantly impact the buying behavior of future tourists who read customer reviews before deciding their travel options.

1.3.4 Innovative Internet-Enabled Business Models

The growing popularity of the Internet, smartphones, and digital media provides opportunities for a company to not only use fragmented new-age marketing effectively to promote existing products, but also to come up with innovative business models where product demo, customer acquisition, and order fulfillment can also take place online.

Innovative business models may include the following:

- **Online Marketplaces**—Several e-commerce companies have created global online marketplaces for selling books, consumer goods, and other products. In such business models, customer acquisition is usually initiated through the company's website. The company coordinates with its multiple suppliers to source products; samples, demos, and product reviews are provided on the website; customers make their purchases online; and items are shipped directly to customers.

Examples of Online Marketplaces:

- Book publishing and retail businesses, which historically gained much success using traditional business models, have been significantly affected by the advent of online marketplaces such as Amazon, eBay, Alibaba, and Flipkart.
- Online payment processors such as PayPal, Stripe, Braintree, and Google Wallet have simplified the way in which payments can be processed by businesses, and have enabled even small start-up companies to sell their products globally.

- **Online Services**—Online services have significantly impacted many traditional product and service industries by transforming existing business models and creating new ways to conduct business.

Examples of Online Services:

- Global Positioning Systems (GPS) and online maps have made physical maps redundant.
- The gaming industry has transformed predominantly to the online community with options for participants to play against opponents from various locations.

- **Online Networking**—The Internet has made the world a smaller place. People can now have access to their networks at all times. These changes have significantly impacted the way in which people communicate with each other and, in turn, have created new possibilities for innovative business models.

Examples of Online Networking:

- Social media channels such as LinkedIn, Twitter, WhatsApp, Facebook, and Google+, have significantly changed the way in which people communicate with each other.
- Online search engines such as Google, Yahoo, and Bing make it easy to find information and locate businesses globally.

- **Business Models Using Smartphones**—Smartphones are Internet-enabled mobile phones that also allow people to have an ongoing connection to the Internet. Since individuals usually carry their smartphones with them, mobile apps are becoming increasingly popular. Innovative business models based on the use of smartphones can disrupt several existing business models—more so in industries that rely on other forms of communications and networking.

Examples of Business Models Using Smartphones:

- Several airlines and travel portals have mobile apps to facilitate the ability to book flight tickets using smartphones.
- Some mobile apps allow users to locate nearby restaurants, read reviews, and also post reviews based on their own experience.

Marketing Research today typically involves both online and offline efforts. Online research makes collection and analysis of data easy and cost effective. Pop-up windows, e-mails, etc. can be used for online surveys. Another source of online data is web analytics, which provides information about visitor profiles, visiting patterns, spending patterns, etc. Based on the data collected, companies try to customize their web pages for each visitor with suggestions such as recommended products, similar products, etc. Online surveys are only completed by consumers with access to the Internet so the results may not be representative of the entire population. If online surveys are conducted on the company's webpage or app then the respondent base is further restricted to their own customers. Offline research is similar to earlier models but companies using disruptive models might have to establish their category before conducting offline research.

Examples of Marketing Research for Innovative Internet-Enabled Business Models:

- XYZ, an online travel portal, uses pop-up windows to collect data about customer satisfaction. A short questionnaire of four to five questions regarding user experience is provided to a customer if he or she agrees to be a part of the survey.
- ABC, an e-commerce company, uses e-mails to invite customers to take part in their quarterly survey. Customers can click on the given link if they want to participate in the survey.

1.3.5 Sales and Marketing as a Continuum

It is important for us to note that the fact that we are in the twenty-first century does not make all the earlier avenues of Sales and Marketing obsolete. Some companies marketing consumer goods continue to spend a significant proportion of their marketing budget on conventional mass media marketing. In some cases a seller's marketplace continues to be the reality for certain commodities that have a limited number of producers, or where the production is highly regulated by the government or controlled by monopolies or duopolies. Similarly, in some regions or countries, traditional marketplaces continue to flourish.

Rather than viewing the changes as completely replacing the earlier practices, Sales and Marketing approaches should be viewed as a continuum where recent innovations can co-exist with earlier practices. It is the responsibility of a company's Sales and Marketing teams to make the strategic decisions that will work best to achieve the desired outcomes, given the reality of the markets and particular consumer preferences.

Sales and Marketing students, who read material on the subject, often find it confusing because authors offer varied perspectives that may be difficult to assimilate and comprehend in the present day. Each author's perspective can also vary depending on when the material was written (i.e., where he or she was on the Sales and Marketing timeline), his or her individual or industry preferences and experiences, and other factors. Conversely, the concepts covered in this Sales and Marketing Body of Knowledge (*SMstudy® Guide)* are not limited to the perspective of any particular author or industry. The *SMstudy® Guide* was developed by VMEdu, Inc., a professional education provider which has educated over 400,000 students world-wide in more than 3,500 companies. The forty plus authors, advisors, and reviewers of this book have worked in multiple marketing environments and geographic regions across an eclectic variety of industries. Thus, the insights provided in this book provide comprehensive detail of the principles and concepts related to Sales and Marketing and specifically to Marketing Research. This book also articulates an action-oriented process approach that can be used by Sales and Marketing practitioners to gain a better understanding of the subject, and then conduct comprehensive and effective Marketing Research that supports both the marketing objectives as set out in the Marketing Strategy and the business goals as established in the Corporate Strategy.

1.4 Corporate Strategy Overview

Corporate Strategy is the overall direction of the company (as defined by senior management) that takes into consideration an assessment of the existing capabilities of the company and external opportunities and threats. Corporate Strategy usually coincides with the immediate future fiscal period or it could be developed with a longer-term view (e.g., a three-year plan). It is important to understand the overall Corporate Strategy and its relationship to all areas of the business in order to ensure that activities at all levels are aligned and aimed at achieving overall corporate goals. It is best to have a clear understanding of where the company plans to be in the near and far future, so that the Marketing Research projects can be planned and executed to contribute to the marketing objectives and the overall goals of the business.

Corporate Strategy is a combination of the following:

1. **Senior Management Direction and Insights**—This is provided by senior management based on their experience and insights related to the business.

2. **Corporate Product Strategy**—This defines the products or services the company offers, and the research and development (R&D) efforts required to create them.

3. **Corporate Marketing Strategy**—This defines how the company plans to target, position, market, and sell the planned products, and defines metrics, targets, and budgets for all marketing activities.

4. **Corporate Operations Strategy**—This defines how the company will manage operational activities, manufacture its products (or provide services), and provide the corresponding customer support and warranty.

5. **Corporate Finance Strategy**—This defines how the company will manage its finances, attain funding, and financially sustain its operations. The Finance Strategy should include forecasts and projections and summarize costs, income, and investments.
6. **Corporate Human Resource Strategy**—This maps the human resource capabilities within the company and considers talent management and acquisition needs to sustain growth.

Typically, companies have existing documentation regarding their Corporate Product Strategy, Corporate Marketing Strategy, Corporate Operations Strategy, Corporate Finance Strategy, and Corporate Human Resource Strategy; these must be considered in an integrated manner to define a coherent Corporate Strategy. The level and complexity of documentation for these strategies may vary depending on the size of the company and the breadth of its product portfolio and geographic reach. If formal documentation of these strategies is not available (e.g., as with a start-up company), the teams involved in strategic planning should consider the various strategies using the *SMstudy® Guide* framework and decide on an overall Corporate Strategy, which can then become a benchmark to execute future plans.

Corporate Strategy can be further divided into lower level strategies depending on the complexity of the organization. For example, the Corporate Strategy for an entire company can be divided into strategies for each business unit or geographic region (e.g., country, state, or city), and then subdivided further into a Product or Brand Strategy for each product or brand in a business unit or geographic region. The Product or Brand Strategy is the lowest level in this hierarchy.

Figure 1-3 illustrates the relationship between Corporate Strategy, Business Unit/Geographic Strategy, and Product/Brand Strategy.

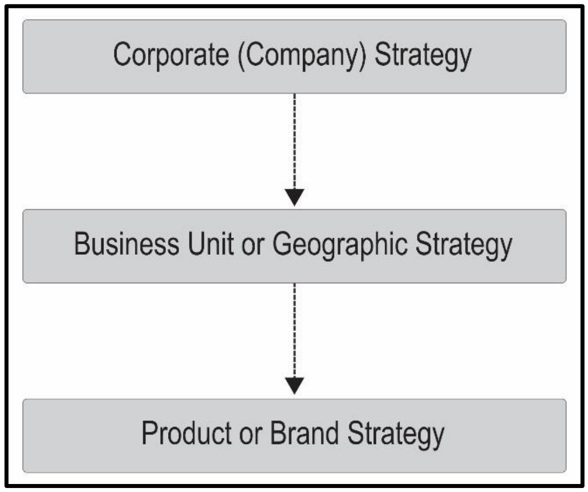


Figure 1-3: Levels of Strategy

While each of the various strategies established in an organization has its own goals and expectations, it is important to note that all activities must align in order to ensure that teams are focused on achieving targets that will contribute to the overall business goals. Alignment of goals across brands, functional areas, and business units, contributes to the attainment of overall marketing objectives and ultimately assists the business in the successful execution of the Corporate Strategy and, therefore, the achievement of business goals. Additional information about Corporate Strategy is described in detail in SMstudy® Guide-Book One, Marketing Strategy (Appendix A).

1.5 Aspects of Sales and Marketing

The *SMstudy® Guide* describes six Aspects of Sales and Marketing as follows:

1. Marketing Strategy (MS)
2. Marketing Research (MR)
3. Digital Marketing (DM)
4. Corporate Sales (CS)
5. Branding and Advertising (BA)
6. Retail Marketing (RM)

Since the *SMstudy® Guide* is geared towards Sales and Marketing professionals or those who desire to work in this field, the six Aspects are based on the six most common and often distinct career fields related to Sales and Marketing. Figure 1-4 illustrates the six Aspects of Sales and Marketing and how they interact with each other.

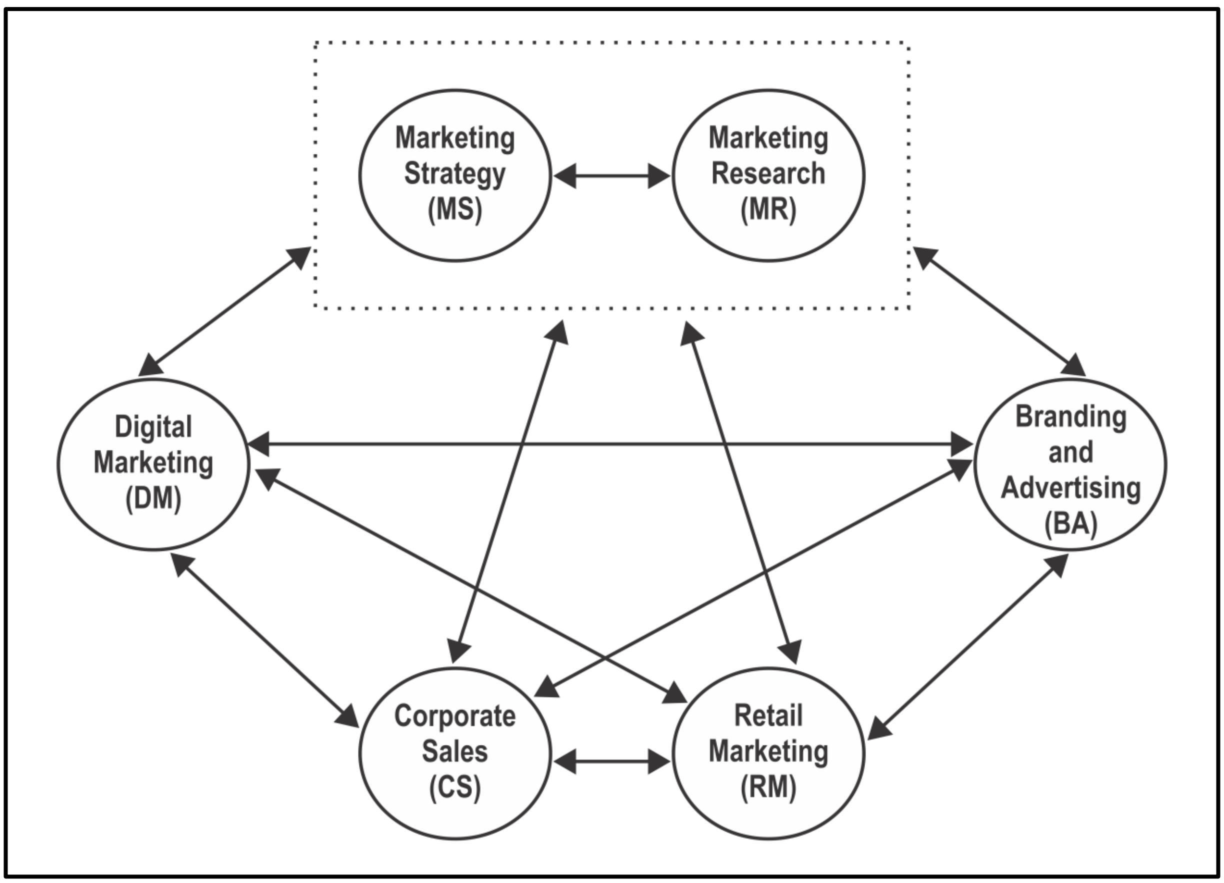


Figure 1-4: Aspects of Sales and Marketing

The two marketing Aspects that are shown in dotted lines at the top of Figure 1-5 (i.e., Marketing Strategy and Marketing Research) are referred to as “Essential Marketing Aspects.” Both of these Aspects are mandatory and should be used to define, measure, and provide direction for the overall marketing efforts of a company.

The four remaining Aspects (i.e., Digital Marketing, Corporate Sales, Branding and Advertising, and Retail Marketing) are referred to as “Optional Marketing Aspects” because one or more of them could be used by a company to reach its marketing goals and, in some instances, not all are applicable. For example, a small company creating phone apps or online games may decide to solely use Digital Marketing; another company manufacturing heavy equipment may use only Corporate Sales; and a large consumer goods company or global fashion chain may decide to use all four Optional Marketing Aspects to reach its marketing goals.

Marketing Strategy (MS), describes how the Aspect of Marketing Strategy aligns with a company’s overall Corporate Strategy and acts as a unifying framework to define and analyze the other Aspects of Sales and Marketing. It also supports the alignment of all marketing resources among all Aspects. Marketing Strategy includes determining internal organizational strengths and weaknesses, as well as external opportunities and threats; identifying and segregating prospective buyers into market segments based on common needs; defining competitive positioning to satisfy specific customer needs; creating pricing and distribution strategies; and defining the metrics, objectives and corresponding budgets for implementation, evaluation, and improvement of all marketing activities.

Marketing Research (MR), which is the focus of this book, explains the concepts of Marketing Research and provides a framework to conduct Marketing Research and to analyze Sales and Marketing data. It also demonstrates how marketing research findings can help the marketing team conceptualize and finalize product features and other components of a company’s Marketing Strategy. In addition, Marketing Research discusses assessment tools that can be used to measure factors that can help drive better corporate decision-making, and in turn more decisive marketing actions. Marketing Research can be conducted for any other Aspect of Sales and Marketing. It is commonly used to test multiple marketing hypotheses in order to better understand consumer behavior, finalize product features, define metrics for measuring marketing efforts, and track and improve marketing activities.

Digital Marketing (DM) includes all marketing activities that use electronic devices connected to the Internet to engage with customers (e.g., computers, tablets, smartphones). These include activities related to creating and managing effective websites and mobile apps as well as promoting a company’s products and brand through various online channels that help meet marketing objectives. Some of the tools pertaining to Digital Marketing include Search Engine Optimization, Search Engine Marketing, Mobile Device Marketing, Social Media Marketing, and E-mail Marketing. This Aspect also demonstrates how an effective Digital Marketing Strategy can be a force multiplier for the other Sales and Marketing Aspects.

Corporate Sales (CS) outlines the best practices and processes to be followed for effective business-to-business (B2B) sales. It provides guidance on activities related to building strong business relationships; successfully working with other businesses to help them see the value in the company’s products and services; understanding procurement management; conducting effective negotiations with other organizations; and ensuring leads generation, qualification, follow-up, and other related activities. It also emphasizes how corporate sales should interface with the other Sales and Marketing Aspects.

Branding and Advertising (BA) includes concepts of product branding, consumer behavior, marketing communication, and public relations. Branding is the process of creating a distinct image of a product or range of products in the customer's mind. This image communicates the promise of value the customer will receive from the product or products. Branding should remain consistent across all channels of communications with the customer. Advertising is defined in the *SMstudy®* Guide as any paid form of non-personal communications to existing and potential customers that promote the company's products through all types of media—such as radio, television, and print. Internet advertising is discussed in the book on Digital Marketing.

Retail Marketing (RM) presents concepts of all marketing activities related to persuading the end customer to purchase a company's products at a physical retail outlet or store, and efficiently managing the supply chain and distribution channels to improve the reach and sales for a company's products. This Aspect also discusses how Retail Marketing interfaces with the other Sales and Marketing Aspects.

1.6 Levels of Sales and Marketing Strategy

The Corporate Marketing Strategy, which is a component of the overall Corporate Strategy, is further divided into various Business Unit or Geographic Strategies, which in turn is further divided into particular Product or Brand Strategies for each product or brand. Figure 1-5 illustrates the relationship between Corporate Marketing Strategy, Business Unit/Geographic Marketing Strategy, and Product/Brand Marketing Strategy.

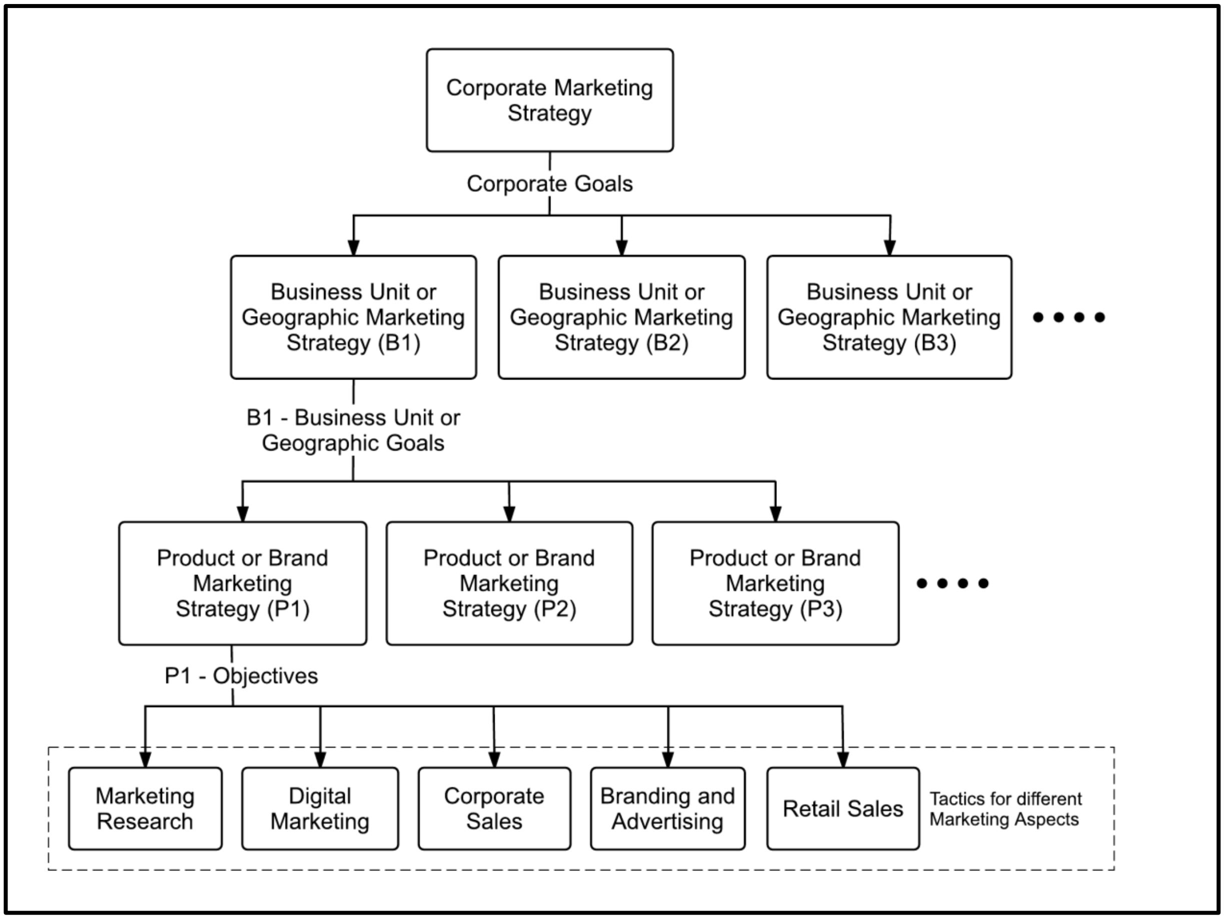


Figure 1-5: Relationship between Different Levels of Sales and Marketing Strategy

The Corporate Marketing Strategy is defined at a corporate level. It defines the overall marketing goals for the company. These general marketing goals drive more specific marketing strategies for each of the company's business units or geographies. Each business unit or geography in turn defines its own goals, which are relevant inputs for each area's particular Product or Brand Marketing Strategies. Each Product or Brand Marketing Strategy (also referred to as 'Marketing Strategy' in the *SMstudy® Guide*) defines Sales and Marketing objectives for each product or brand, which drive specific tactics that align with and often rely on other Marketing Aspects. Marketing Research is one of six Aspects of Marketing defined in the *SMstudy® Guide*, and is the focus of this book. The marketing activities across all Aspects of Marketing are designed with the marketing objectives in mind. Within the strategy for each Aspect, various activities are designed to meet specific targets that the team establishes will provide a measure of success and enable the team to contribute to the overall marketing objectives and, ultimately, to the business goals.

Example of Levels of Marketing Strategy:

Land Development Company

- **Corporate Level:** A land development company wants to grow to be among the top three land development companies in its state.
- **Business Unit/Geographic Level:** The land development company operates two business units: Residential and Retail. A goal of the Residential Business Unit is to grow that unit by 12% within one year; a goal of the Retail Business Unit is to grow that unit by 10% within the same time period.
- **Product/Brand Level:** Within the Residential Business Unit, the company sells three products: Condominiums, Townhomes, and Singles. The Singles Product Marketing Strategy identifies an objective to grow the sale of single units by 15%. To achieve this objective, the teams responsible for building strategy within the various Aspects of Marketing establish specific objectives that are designed to support the overall product objectives and to align with one another.
- **Marketing Aspect Level:** The company's greatest strength is the fact that it is an award-winning leader in 'green' sustainable development. Therefore, the Branding and Advertising team builds specific tactics that incorporate an increase in reach of its messaging around sustainable development. One specific tactic is to leverage billboard and newspaper advertising with the objective of increasing reach of 'green' messaging by 30%. The Digital Marketing team incorporates tactics to support the objective of increasing the 'green' sustainable development messaging, stressing the importance of this trend, and positioning the company as a leader in the industry, through the use of various social media channels. One specific tactic is to leverage blogs and online public relations with the objective of increasing the company's rankings in online searches related to keywords, such as 'sustainable development.' The tactics of each Marketing Aspect are aimed at achieving their own specific objectives; however, both support the overall Singles Product Strategy objective of achieving a 15% growth in sales for this product line.

Example of Levels of Marketing Strategy:

Global Automobile Company

- **Corporate Level:** A global automobile company specializing in manufacturing luxury automobiles has a corporate goal to grow the company by 8% in the upcoming year by launching new models of cars in its existing locations and entering new market segments.
- **Business Unit/Geographic Level:** The automobile company has been organized into multiple business units based on geographies where it conducts business. Each business unit has business unit goals that contribute to the company's overall corporate goals. The business unit goals for next year are 5% growth in the United States, 10% growth in China, 4% growth in the United Kingdom, 12% growth in India, and 6% growth in Germany.
- **Product/Brand Level:** To meet the 10% growth target in China, the marketing team in China plans strategies for the three existing brands in the market (i.e., 'Ceres,' 'Pallas,' and 'Vesta') and also plans to launch a new brand, 'Juno.' Each brand targets a different customer segment. Ceres is an entry-level sedan, targeted at working professionals who aspire to have a luxury car; Pallas is a minivan, primarily targeted at families with children; Vesta is a four-wheel drive sport-utility vehicle for individuals who want both on- and off-road capability and to still be able to go on long drives; the new brand, Juno is a convertible that the Chinese business unit plans to target at young persons who want a stylish and fun car.

  Each brand team creates a Marketing Strategy for its brand. When creating the Marketing Strategy, the team considers the strengths, weaknesses, opportunities, and threats for the brand; defines the market and identifies the different market segments; identifies the brand's competition; finalizes the target market segment for the brand; analyzes the target market to create a differentiated positioning for the brand; and then finalizes the pricing and distribution strategies. Each team then determines the appropriate metrics and objectives that will help reach the team's growth target, and a budget is allocated to each Marketing Aspect. Juno's key metric is sales and its main objective is to sell 25,000 cars in the Chinese market the year after the vehicle is released. The Marketing Strategy team for Juno decides to use Digital Marketing, Branding and Advertising, and Retail Marketing to reach out to its target market segment. Juno's product strategy team sets a budget of $1 million for Digital Marketing to sell 3,000 cars, $10 million for Branding and Advertising to sell 10,000 cars, and $15 million for Retail Marketing to sell 12,000 cars.
- **Marketing Aspect Level:** The metrics, objectives, and budgets allocated to each of the Marketing Aspects become inputs for those Aspects. For example, the Digital Marketing team may decide to create a high-quality website with their budget of $1 million and an objective of selling 3,000 cars.

1.7 Marketing Strategy Overview

All successful products or brands need well-planned marketing strategies in place to ensure that they satisfy the goals set by the corresponding Business Unit or Geographic location, and in turn the overall Corporate Marketing Strategy. Marketing Strategy is therefore one of the most crucial Aspects of Sales and Marketing. It defines a product or brand's unique value proposition, target markets, and the specific strategies to be used to connect with defined audiences. It also specifies the pricing and distribution strategies for a product or brand, and outlines the specific metrics, objectives, and budgets for all its marketing activities.

Among the outputs of Marketing Strategy are the specific Aspects that will be used to achieve the overall marketing objectives. Although Marketing Research is one element of a variety of Aspects of Marketing that businesses use to grow their business and achieve corporate goals, it is considered an "Essential Marketing Aspect," as it is a significant contributor to all other Aspects of Marketing. The findings from all marketing research efforts inform key decisions relating to understanding and defining target markets, selecting channels, determining resources, and selecting and measuring marketing activities, among others. Effective Marketing Research will result in customer-focused Sales and Marketing planning across all Aspects of Marketing. For a comprehensive understanding of how to build and execute a Marketing Strategy that aligns all Aspects of Marketing toward achieving both the marketing objectives and the business goals, consult the SMstudy® Guide-Book One, Marketing Strategy.

1.8 Marketing Research Overview

Marketing Research is one of several Aspects of Marketing discussed in the *SMStudy® Guide* series. Other Aspects discussed in this series include Digital Marketing (DM), Corporate Sales (CS), Branding and Advertising (BA), and Retail Marketing (RM). The strategies for each of the Aspects of Marketing are derived from the outputs of the Marketing Strategy (MS), which are described in the *SMStudy® Guide* series on Marketing Strategy.

Marketing Research is the systematic process of collecting, processing, and analyzing data to provide required information to decision makers. Marketing Research is linked to all other Aspects of Marketing as it provides critical insights that inform key decisions in all other marketing planning and strategies. For example, it can provide information for pricing a new product or for designing a new Mobile App or for finalizing the new TV advertisement for a product.

The processes associated with planning and executing a marketing research project include understanding the research problem; deciding a suitable research design; collecting, processing, and analyzing data required to solve the problem; interpreting data; and presenting the findings/recommendations of the research project to the key decision makers.

1.8.1 Define Research Problem and Choose Research Design

Chapter two of this book focuses on defining the research problem and choosing a research design. A marketing research project starts with defining the research problem and ends with the solution to the problem. Since this step sets the course of the research project, defining the research problem is a critical step in the lifecycle of the research project. Once the research problem is properly defined, the next step is to choose a research design that can address the problem and objectives. Different research methods are used for different types of research projects based on the types of information needed. Further, every research project is unique in itself; therefore, there is no single research design best suited for a particular type of research. Many alternative research designs can satisfy the same research objectives, and the researcher needs to select a design that will be most suitable for the current project.

Figure 1-6 provides an overview of the processes discussed in Chapter 2, Define Research Problem and Choose Research Design.

2.1 Define Research Problem

INPUTS
1. Management Problem*
2. Existing Marketing Strategy
3. Corporate Strategy*
4. Senior Management Direction and Insights

TOOLS
1. Use Cases
2. Fishbone
3. 2x2 (Two-by-Two) Matrices
4. Meetings and Discussions*
5. Situation Analysis*
6. Symptomatic Situation Analysis

OUTPUTS
1. Research Problem and Objectives*

2.2 Choose Research Design

INPUTS
1. Research Problem and Objectives*
2. Senior Management Direction and Insights*
3. Expert Judgment
4. Background Information*
5. Available Information

TOOLS
1. Meetings and Discussions*
2. Available Information Evaluation
3. Exploratory Research Design*
4. Descriptive Research Design*
5. Causal Research Design*

OUTPUTS
1. Selected Research Design*

Figure 1-6: Define Research Problem and Choose Research Design Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

1.8.2 Data Collection

The third chapter of this book discusses data collection methods for two types of research data—primary data and secondary data. Many significant marketing decisions are made based on the analysis of the data collected from a research project. Hence, the quality and relevance of the data is very important for any marketing research project.

Other important factors to consider are the availability of data and the affordability of the process required to collect it. The researcher needs to decide whether to collect primary data or spend the research budget exclusively on secondary data. Researchers usually prefer to examine the utility of low-cost and readily available secondary data first to see whether they can partly or fully solve the research problem under investigation without collecting costly primary data. The source of the secondary data can be internal or external. The sources may include books or periodicals, published reports, data services, and computer data banks. When the needed data is non-existent, outdated, incorrect, or inadequate, the researcher needs to collect primary data. Most marketing research projects do include some primary data collection. Primary data may be obtained from consumers, subject matter experts, random samplings of a target segment, organizations, and other sources.

Figure 1-7 provides an overview of the processes discussed in Chapter 3, Data Collection.

3.1 Collect Secondary Data	3.2 Collect Primary Data
INPUTS 1. Senior Management Direction and Insights 2. Research Problem and Objectives* 3. Selected Research Design* **TOOLS** 1. Meetings and Discussions* 2. Internal Reports* 3. Annual Reports 4. Government Publications* 5. Commercial Sources 6. General Media* 7. Bibliographic Database **OUTPUTS** 1. Collected Secondary Data*	**INPUTS** 1. Research Problem and Objectives* 2. Selected Research Design* 3. Senior Management Direction and Insights 4. Secondary Data* 5. Sampling Design* **TOOLS** 1. Observation Techniques* 2. Experiments* 3. Qualitative Techniques* 4. Quantitative Techniques* **OUTPUTS** 1. Collected Primary Data*

Figure 1-7: Data Collection Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

1.8.3 Data Processing and Data Analysis

Chapter four of Marketing Research takes a close look at techniques and methods that convert data into useful information that can answer the marketing research question. Two processes are explored in this chapter—Data Processing and Data Analysis. Data Processing is used to organize, manipulate, and transform raw data into useful information that is ready for analysis. Data Analysis is used to transform processed data into intelligence which helps in decision-making.

It is important to have a structured framework for data processing and analysis. Hence, the researcher establishes, in the creation of the research design process, how much data will be collected and how the data will be processed and analyzed. Establishing a plan for processing and analyzing data before collection also helps ensure that only required information is collected, that the data is formatted in a standardized way for easy processing, and that the output of the analysis will satisfy the research objectives and solve the research problem.

Figure 1-8 provides an overview of the two processes associated with Data Processing and Data Analysis.

4.1 Data Processing

INPUTS
1. Primary Data*
2. Secondary Data*

TOOLS
1. Coding*
2. Data Cleaning*
3. Weighting
4. Variable Respecification
5. Scale Transformation
6. Tabulation*
7. Computer Processing*
8. Data Mining*

OUTPUTS
1. Processed Data*

4.2 Data Analysis

INPUTS
1. Research Problem and Objectives*
2. Data Analysis Objectives*
3. Selected Research Design*
4. Processed Data*
5. Expert Judgment

TOOLS
1. Statistical Inference*
2. Bivariate Data Analysis
3. Multivariate Data Analysis
4. Statistical Packages*

OUTPUTS
1. Analyzed Data*

Figure 1-8: Data Processing and Data Analysis Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

1.8.4 Data Interpretation and Reporting

Chapter five of this book focuses on interpreting and reporting analyzed data. This is an important step in transforming the analyzed data into meaningful and reliable information that can be used to solve the research problem and ultimately inform key marketing and business decisions. In Data Interpretation, all of the possible interpretations from the analyzed data are determined. Interpreting the data appropriately, accurately, and from the right perspective is important to ensure that sound marketing decisions are made based on the interpretations. Reporting is the final process of a Marketing Research project. In this process guidelines are provided to create the Marketing Research report based on the data interpretation and the research problem and objectives. Reports and presentations have to be created with the target audience in mind. As this is typically the only component of the study that management and key decision makers are exposed to, the overall evaluation of the research project rests almost completely on how well this information is communicated.

The two processes discussed in this chapter are shown in figure 1-9.

5.1 Data Interpretation

INPUTS
1. Analyzed Data*
2. Research Problem and Objectives*

TOOLS
1. Tables*
2. Charts*
3. Expert Judgment

OUTPUTS
1. Interpretations*

5.2 Reporting

INPUTS
1. Interpretations*
2. Research Problem and Objectives*
3. Senior Management Direction and Insights

TOOLS
1. Oral Reporting
2. Report Writing*
3. Report Format
4. Presentation Software

OUTPUTS
1. Research Report*
2. Recommendations

Figure 1-9: Data Interpretation and Reporting Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

2. DEFINE RESEARCH PROBLEM AND CHOOSE RESEARCH DESIGN

When a marketing decision maker is faced with a management problem, he or she may decide to use marketing research in order to arrive at an objective, informed decision, or course of action. In marketing research, the first step is to define the research problem and objectives of the research. Once the research problem is properly defined, the next step is to choose a research design that can address the problem and objectives. Different research methods are used for different types of research projects based on the types of information needed. Further, every research project is unique in itself; therefore, there is no single research design best suited for a particular type of research. Many alternative research designs can satisfy the same research objectives, and the researcher needs to select a design that will be the most suitable for the current project.

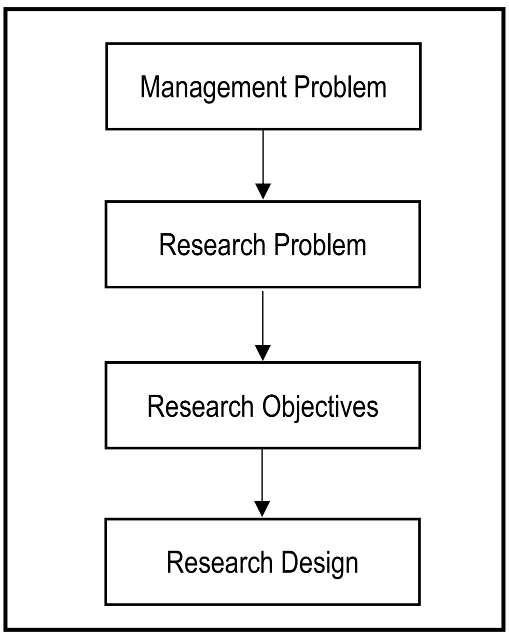


Figure 2-1: Flowchart Depicting Initial Stage of Marketing Research

Figure 2-2 provides an overview of the two processes associated with this initial stage of a research project.

2.1 Define Research Problem—In this process, the researcher understands the situation at hand and defines the research problem, which in turn helps to determine the objectives of the research.

2.2 Choose Research Design—In this process, the researcher examines the background information, consults with relevant parties and experts, and finalizes a research design that will be most suitable to fulfill the research objectives within the given conditions.

2.1 Define Research Problem

INPUTS
1. Management Problem*
2. Existing Marketing Strategy
3. Corporate Strategy*
4. Senior Management Direction and Insights

TOOLS
1. Use Cases
2. Fishbone
3. 2x2 (Two-by-Two) Matrices
4. Meetings and Discussions*
5. Situation Analysis*
6. Symptomatic Situation Analysis

OUTPUTS
1. Research Problem and Objectives*

2.2 Choose Research Design

INPUTS
1. Research Problem and Objectives*
2. Senior Management Direction and Insights*
3. Expert Judgment
4. Background Information*
5. Available Information

TOOLS
1. Meetings and Discussions*
2. Available Information Evaluation
3. Exploratory Research Design*
4. Descriptive Research Design*
5. Causal Research Design*

OUTPUTS
1. Selected Research Design*

Figure 2-2: Define Research Problem and Choose Research Design Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

2.1 Define Research Problem

A marketing research project starts with defining the research problem and ends with the solution to the problem. Since this step sets the course of the research project, defining the research problem is the most important step in the lifecycle of the research project. The success of the research project is, in fact, highly dependent on whether the problem statement has been clearly and accurately identified at the outset. The research problem provides the context for the research design, and therefore should be focused, appropriate, and clear in order to effectively guide the efforts of the research team throughout the project lifecycle.

Defining the research problem involves analysis of the management problem along with insights into the company's future direction and strategy. A comprehensive analysis of many factors helps the team define a proper research problem and set clear research objectives. There are numerous examples in which research problems are defined based on intuition, and this approach can lead to the failure of the marketing research project. Clear definition of the marketing research problem will help ensure that useful and meaningful information is gained from the research project and will help the team avoid additional costs and delays throughout the research project.

Identification of the research problem is also one of the most difficult phases in a project. Often the problem or the source of the problem is too broad or vague for the researcher to investigate. The problem faced by the decision makers needs to be analyzed in order to narrow the broad management problem area into a research problem. The research problem definition should be neither too broad nor too narrow. A research problem which is too broad does not provide enough focus for subsequent steps (e.g., "Improving the image of a brand" is too broad of a problem area to investigate). If the problem definition is too narrow, there is the possibility of overlooking or missing important components of the problem (e.g., if the investigation of low sales is limited only to analyzing customer reactions to a price increase, then the research team would likely miss other factors such as seasonality and point-of-sale promotions).

During this process, the researcher tries to understand the background of the management problem, gathers relevant information and context, has discussions with important stakeholders, and conducts an analysis.

Figure 2-3 shows the inputs, tools, and outputs for the *Define Research Problem* process.

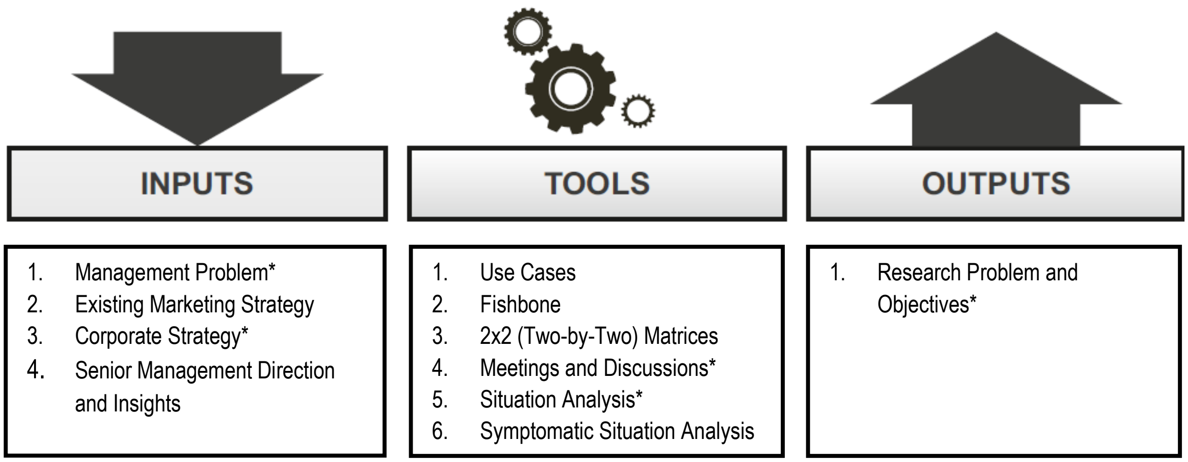


Figure 2-3: Define Research Problem—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

2.1.1 Inputs

2.1.1.1 Management Problem*

Marketing research is undertaken to gather information that is partially or fully unavailable. The information may be an evaluation of product performance, the study of an industry trend, or the analysis of a market selected for the launch of a new product. The outcome of the research should help management, or the sponsor of the research assignment, make decisions concerning any kind of business requirement. The primary objective of marketing research is to provide management with enough data or evidence to make decisions on issues objectively and reduce the risk of a business failure.

The management decision problem or management problem states the business-related issues, whereas the marketing research problem states the information that is required to solve the problem. For example, a sales manager wanting to increase market share for product X is a management problem. One of the research problems for this management problem could be “how can we encourage the sales team to generate more sales?” The research project will provide information that will help the sales manager devise an attractive incentive structure to motivate the sales people, which in turn will solve the management problem of increasing market share.

The researcher needs to understand and analyze the management problem and translate it into a research problem that marketing research can address.

Examples of Management Problems:

- A large clothing retailer operating in the UK wants to open franchise stores in North America. One management problem might be “How can we increase brand presence in the new market?” One of the research problems for this management problem could be, “How is the brand performing currently against the competition in other markets?”
- An established coffee shop is concerned with increasing competition from national franchises. Its management problem is “to improve or sustain market share.” One of the research problems could be, “What features are important to customers when choosing a coffee shop?”

2.1.1.2 Existing Marketing Strategy

Execution of the existing marketing strategy requires information such as market size, demand, customer behavior, existing competition, and future forecasts. Marketing research can be performed frequently since every business runs in a dynamic environment.

Changes in technology over the years, consumer buying trends, economic conditions, and changes in business strategy are a few examples that may require marketing research for management when it is strategizing future business plans.

When marketing research is conducted for a research problem concerning an existing offering, the researcher needs to examine the existing marketing strategy for that offering when defining the research problem. Examining the existing marketing strategy will help the researcher understand the market-oriented strategies of the product and other important aspects of the offering.

Example of Existing Marketing Strategy:

- A leading fast food chain entering a new market/geography implements the existing marketing strategy of understanding popular local flavors and incorporating them into its menu to attract its target group in the new market. Marketing research goes hand-in-hand at every step of decision-making in this marketing strategy.

  The research problem in this case is identifying local flavors that can be incorporated into the menu. The marketing research starts with understanding the eating habits and consumer behavior of the target group to develop the menu for each kind of meal. This is followed by testing—various fast-food items that incorporate local flavors as well as some standard menu items are introduced to a limited market. The final food items are selected and placed on the menu.

  Here, marketing research helps the company enter a new geography with a marketing strategy based on a robust analysis of the market. The next steps in marketing are again supported by research for testing the advertising creatives, understanding the image of the brand in the market, designing the best media plan for the category, and understanding the impact of the marketing activities on consumers and, consequently, on the brand.

2.1.1.3 Corporate Strategy*

Understanding the Corporate Strategy is essential in identifying the problem the marketing research needs to address. Corporate Strategy, which is defined by the company's senior management team, provides overall direction for the company and helps assess the existing capabilities of the company as well as external opportunities and threats. Existing documentation on Corporate Strategy—a combination of Product Strategy, Marketing Strategy, Operations Strategy, Finance Strategy, and Human Resource Strategy—needs to be considered because it applies to the context of the problem when determining what information is needed, and how that information will be used in decision-making. Figure 2-4 shows the various strategies that make up the company's corporate strategy.

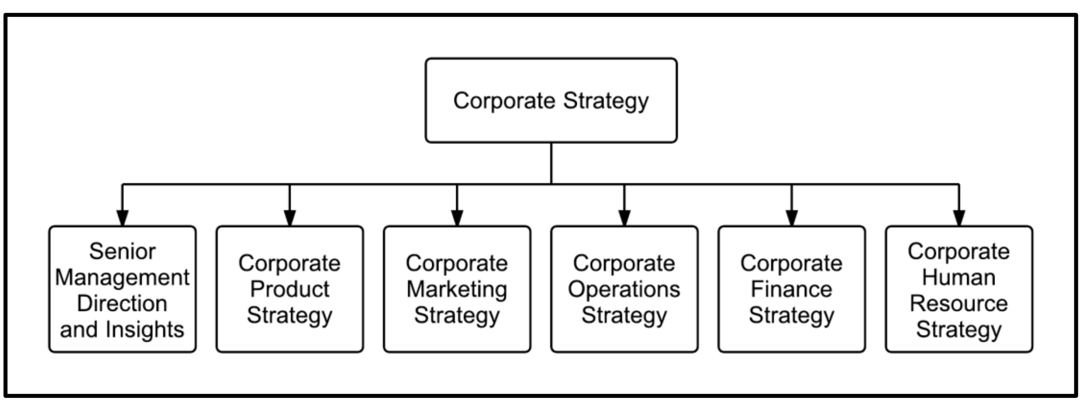


Figure 2-4: Corporate Strategy Overview

Marketing Research is often used to define Product Strategy for future markets by analyzing consumer behavior. The existing Product Strategy, along with an external view of competition and possible innovation opportunities, provides valuable insights while defining a market research problem related to entering a new market or launching new product variants. For example, the existing Product Strategy of a fast-food product will have information on multiple taste variants of the product. This information will be helpful in determining the marketing research problem for inclusion in a new market or introduction of a new variant.

The information on target segments, defined in the company's Marketing Strategy, will help marketing researchers create customer personas for each segment. In the mobile phone industry, for example, users who prefer advanced features such as e-mail and Internet browsing have different personas than users who mainly care about signal strength and basic voice. (There may, of course, be overlaps—all types of customers want longer battery life.) Without a comprehensive review of target segments, a marketing researcher may be analyzing the problem among users who are not associated with the product.

Information from the existing Operations Strategy will provide inputs on defining problems related to operational processes that need to be identified through marketing research. For example, in a low-priced market segment, firms need efficient large volume operations to be more profitable and have a competitive advantage. Department stores and other high-margin firms often set the selling price of items at the manufacturer's suggested selling price. The large volume operators that enjoy the economies of scale obtain a higher profit margin by receiving a lower wholesale price. Mega-discount stores such as Walmart are more likely to pass the savings from the lower costs on to customers in the form of lower prices, thus maintaining their competitive advantage.

Example of Corporate Strategy:

- A classifieds website focused on selling used goods changed its corporate strategy to start selling new products from well-known brands along with providing a meeting place for online buyers and sellers. Thus, the target groups of the marketing research for understanding consumer behavior have to be modified accordingly to include those who are interested in buying new products online as well as those who prefer to buy only used products online.

  In this way, the corporate strategy directs the marketing research and ensures that its results are useful for refining the corporate strategy and creating marketing tactics in line with the needs of the right set of consumers.

2.1.1.4 Senior Management Direction and Insights

Organizations require leadership across different business units and functions in order to implement best practices while striving to carry on business as usual. Most organizations have both strategy and leadership teams integrated in their senior management team. This senior management team comprises members with vast experience in their respective fields who provide useful insights and strategic thinking to functional teams. They also help the organization position itself for its long-, medium-, and short-term goals.

Since the senior management team regularly receives information from the functional strategy teams, it acquires a comprehensive view of the company, and as a result can provide insights into problems and their causes, even before they are articulated through situation analysis and market opportunity analysis. These insights allow the senior management team to carefully define the problem, which individual functional teams may not be able to do. For example, in a retail shoe chain, the manager may decide that the product-positioning and inaccurate floor set-up are the causes for a drop in sales. However, inputs from senior management may prove that the actual cause is improper training for sales staff regarding quality customer service.

Given their industry knowledge and long-term view of the company's goals, members of senior management may help identify problems related to entering new markets. Without the involvement of senior management and its leadership, insights, and vision, the marketing research team risks overlooking the strengths and weaknesses in other functional areas, defining a problem that does not complement the corporate strategy, and/or moving in a direction that is not aligned with corporate objectives.

Example of Senior Management Direction and Insights:

- An auto-parts manufacturer commissioned a marketing research study to understand need gaps in consumer categories. According to the research proposal, the study would be conducted within two segments—four-wheeler owners and heavy vehicle drivers. After this proposal was presented to senior management, the target groups of the study were modified. The first segment was changed to four-wheeler owners and four-wheeler mechanics, instead of only owners, and the second segment was changed to heavy vehicle fleet owners and managers, instead of drivers.

  Senior management had insights into the industry that marketing did not, and this knowledge affected the parameters of the study. The decision to select an auto-part is most often made jointly by four-wheeler owners and mechanics and for heavy vehicles is decided by the fleet owner or manager for all the vehicles in the fleet.

  Thus, senior management direction and insights are necessary to refine the research problem and the research scope before it is implemented.

2.1.2 Tools

2.1.2.1 Use Cases

A use case is a methodology or list of steps that defines the interactions between stakeholders and a system. A use case is an effective tool for identifying a research problem. Use cases can be used to describe how a customer's needs are met (or not met) by the product or service offering. A use case enables marketers to examine the business problem that they are trying to solve and frame the choices that will be examined through marketing research. The steps for generating a use case are as follows:

1. Identify the customer (or actor).
2. Identify other stakeholders (other actors in the system).
3. Identify the interactions they will undertake in the system (the uses).
4. Describe the actions and interactions between the actors and the environment (system). Actions are captured as arrows between stages in the journey.

For example, a bank might look at how an individual consumer responds to the use of an ATM when they are looking to do a transaction (i.e., deposit, withdraw, make a payment, buy an investment). Looking at each of these transactions, the use case helps to visualize customer interactions in order to improve or change the customer experience. This information provides context for crafting the marketing research problem statement. By exploring the customer experience and where it may or may not meet certain customer needs (use cases) a research study can be framed more effectively. Figure 2-5 shows a use case for customer interaction with an ATM.

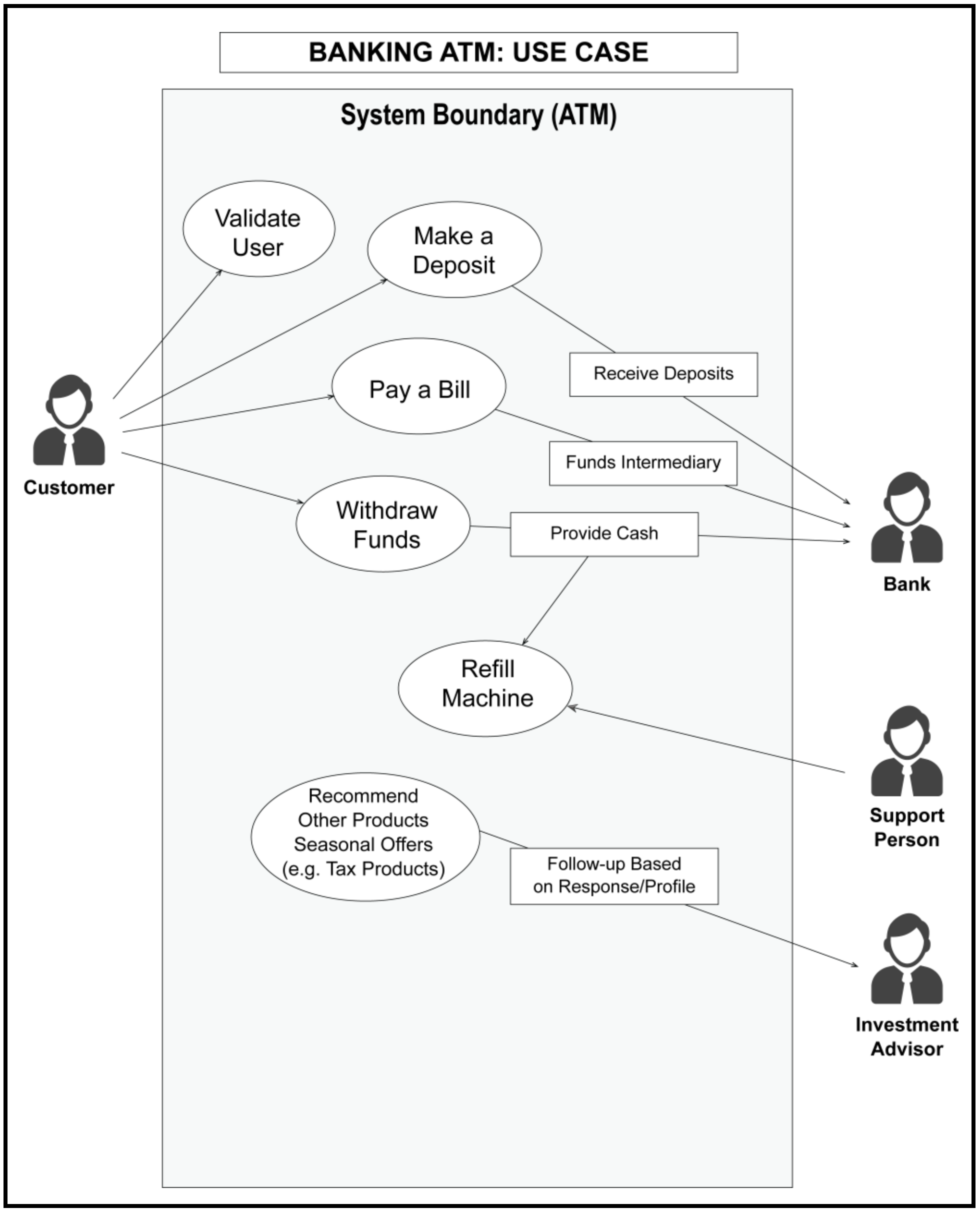


Figure 2-5: Use Case Example

2.1.2.2 Fishbone

A Fishbone diagram, also called a cause and effect diagram, is a team tool for identifying possible causes to a problem and is often used in quality assurance measurement. The steps for creating a fishbone diagram are as follows:

1. Identify the hypothesis problem statement (draw a box around this sentence and then a horizontal line running from the box across the page or whiteboard).

2. Through a brainstorming exercise, identify major common categories that could be the cause of the "problem."
3. If brainstorming is not producing agreed upon categories, use generic terms, such as methods, processes, measurements, people, machines, or environment.
4. Write each of the categories as branches off of the main arrow (the horizontal line running from the hypothesis problem statement).
5. Ask "Why does this happen?" for each of the causes and continue to dive deeper with each round of questions to generate a better understanding of where the problem originates.

For example, a bank might look at the following hypothesis: customers cannot easily buy investments. The research team could use a fishbone to breakdown the elements of the problem to explore how each of these factors creates an impact on the outcome. Identifying elements of the customer experience and how the people, the information, the environment (configuration), the furniture, the equipment, and the methods impact what the customer must do to purchase an investment can help the bank determine the root of the problem. By extension, the fishbone diagram can help identify where further data and research is required to create a more expedient and satisfactory purchase process for the customer. Figure 2-6 shows a fishbone for this problem.

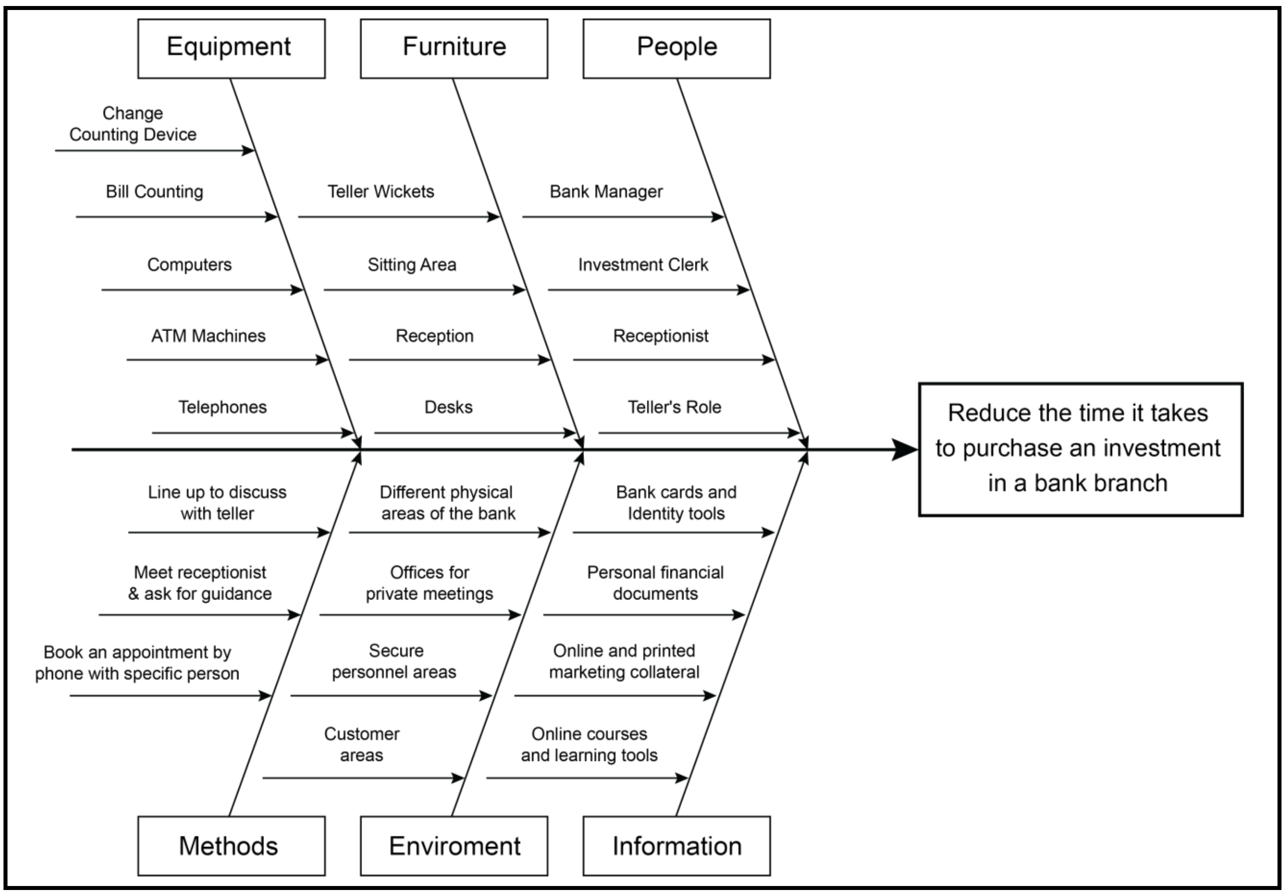


Figure 2-6: Fishbone Example

2.1.2.3 2X2 (Two-by-Two) Matrices

The use of a 2X2 matrix for improving decision-making and problem identification crosses many fields of study and applications from game theory to consulting. A 2X2 matrix can expand options and force the team to think muti-dimensionally. Rather than seeking one answer to a problem the team looks at relationships, considering the impact that multiple adjustments can have on the potential solution. It creates focus around the problem statement to allow exploration of various types of opposition: direct, complementary, and reflexive. The choice of axes indicates the type of opposition being explored.

- **Direct Opposition**—In a 2X2 matrix that uses direct opposition, the team examines opposing options from which to choose—i.e., selecting one option from options that are distinct from one another. Stephen Covey's Time Management Grid, used for prioritizing work on the basis of urgency and importance, is an example of a 2X2 Matrix based on direct opposition. Figure 2-7 shows a Time Management Grid[1]

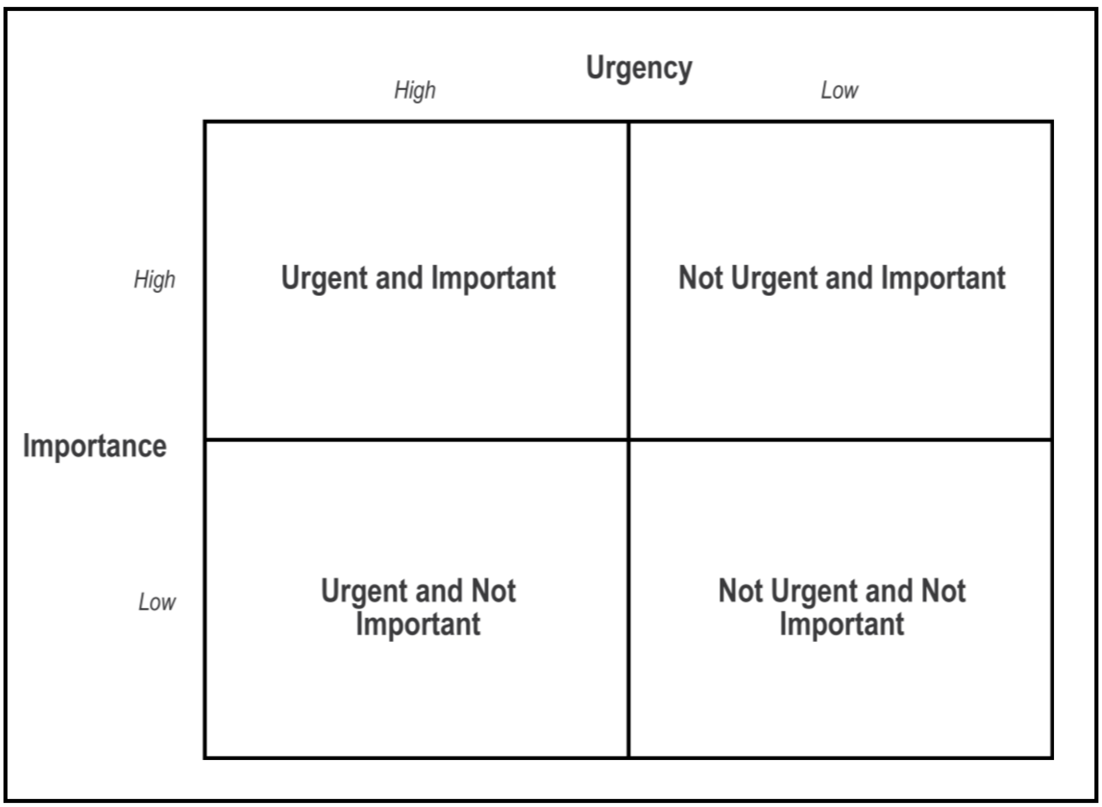


Figure 2-7: Time Management Grid

[1] Stephen R.Covey (1989). "The 7 Habits of Highly Effective People". Free Press

- **Complementary Opposition**—In a 2X2 matrix that uses complementary opposition, qualitatively different options that can be interdependent are explored. BCG matrix, given in Figure 2-8, explores complementary opposition. It analyzes product portfolio on the basis of market share and market growth rate. A thorough discussion of the BCG Matrix is provided in SMstudy® Guide-Book One, Marketing Strategy (section 2.1).

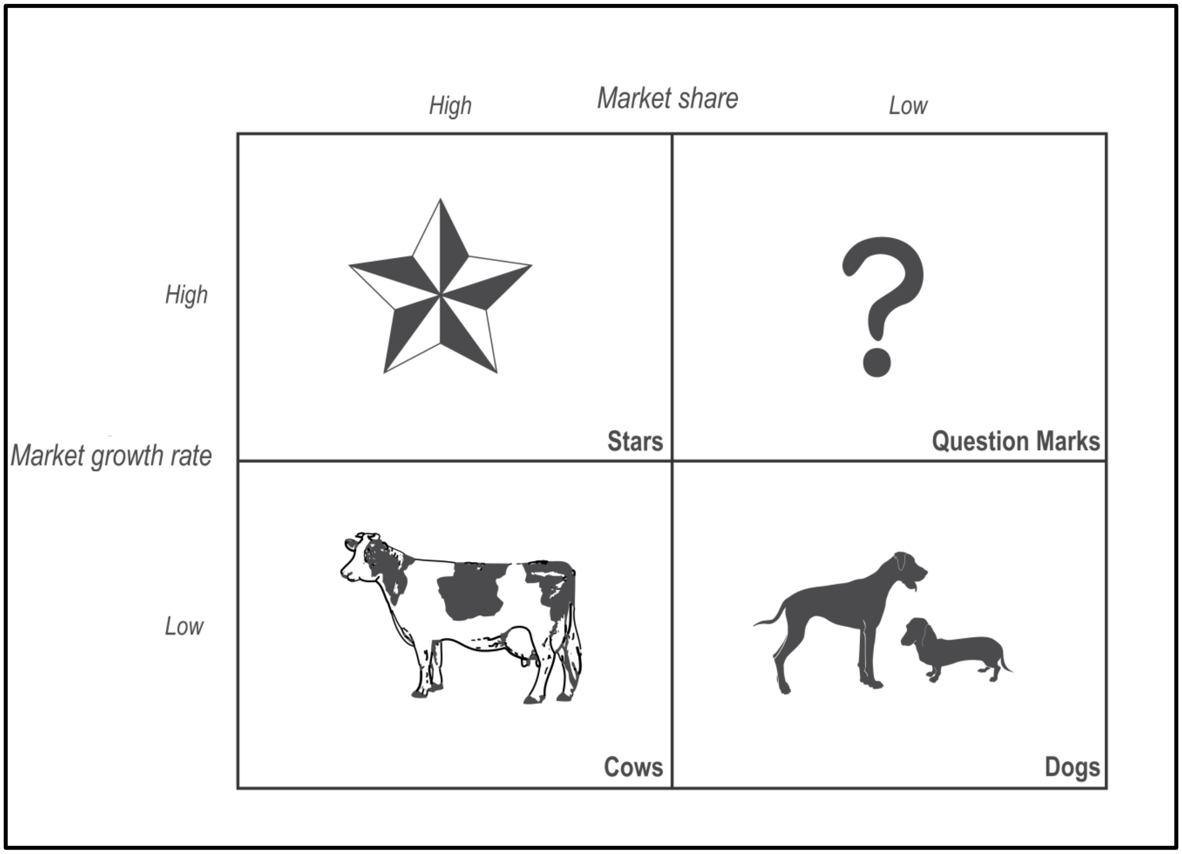


Figure 2-8: BCG Matrix

- **Reflexive Opposition**—In a 2X2 matrix that uses reflexive opposition, options in the same category are used on both axes (but they can be viewed from more than one perspective). Johari Window[2], which looks at knowledge from different perspectives- known to self and known to others, is an example of a 2X2 Matrix based on reflexive opposition. Figure 2-9 shows a Johari Window.

[2] Luft, J.; Ingham, H. (1955). "The Johari Window, A Graphic Model of Interpersonal Awareness". Proceedings of the Western Training Laboratory in Group Development, University of California

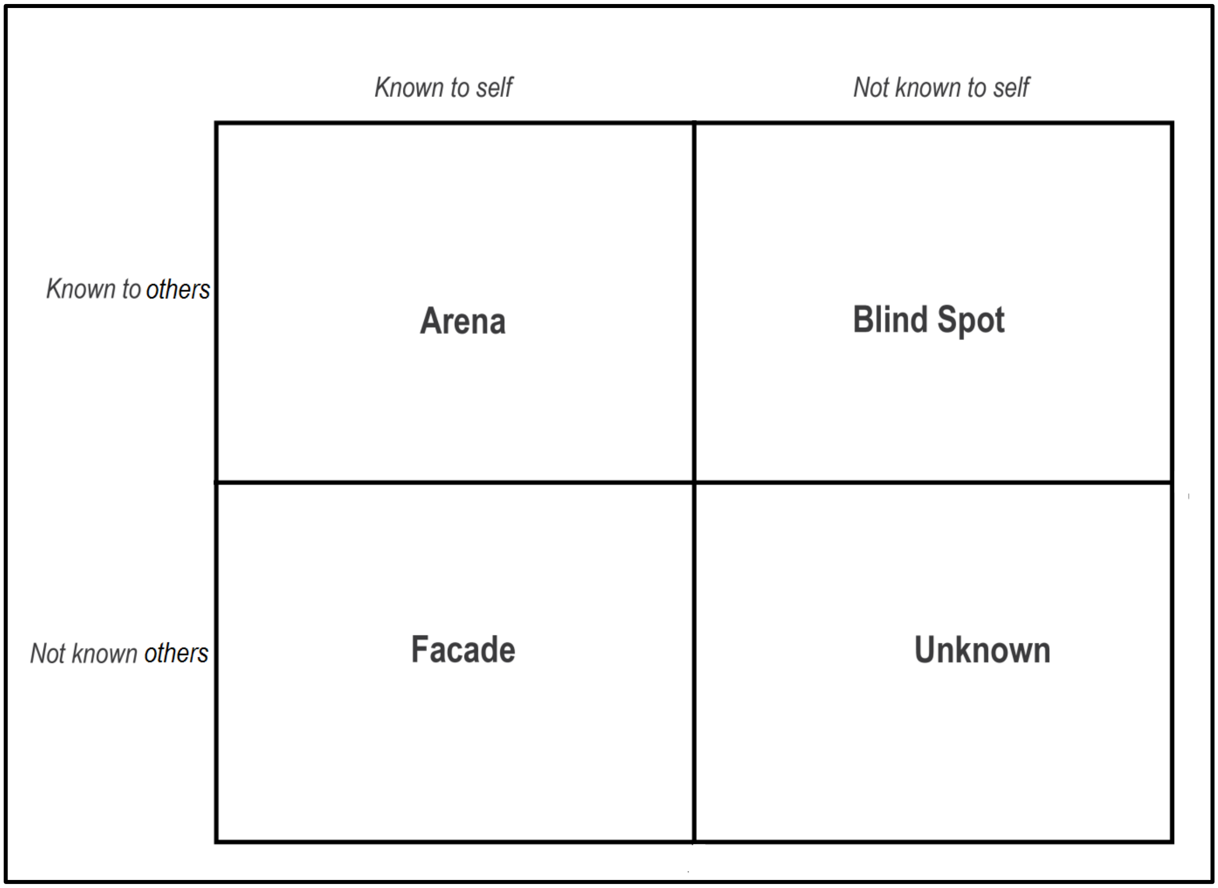


Figure 2-9: Johari Window

The steps for creating a 2X2 matrix are as follows:

1. Identify the hypothetical problem statement, acknowledging the problem that is resisting a solution.
2. Envision the outcome you are trying to generate.
3. List the features and characteristics of the problem, identifying fears, hopes, values, and essential elements of the problem statement.
4. Organize the features into themes.
5. Prioritize the themes.
6. Draw the 2X2 matrix and test the prioritized themes against each other, closing in on a pair.
7. Name the four quadrants to provide guidance on the placement of the content.

Test for coverage to ensure that the themes and quadrants cover the core elements of the problem statement definition.

2.1.2.4 Meetings and Discussions*

Detailed knowledge of the management problem is extremely important and helps a researcher determine whether a marketing research project is actually needed and, if needed, what the research problem statement and goal(s) of the research should be. Clear communication between the researcher and the management problem owner in the problem definition stage is the most important factor for improving the usefulness of research. The researcher needs to be involved in detailed meetings and discussions with the management problem owner or the manager/decision maker. The researcher gathers all relevant information and data regarding the management problem and asks as many questions as possible during discussions with the managers. To understand the background of a marketing research problem, the researcher must understand the client's firm and industry as well as the management problem itself. During discussions with managers and experts, the researcher should ask questions regarding, but not limited to the following:

- the history, current status, and future objectives (The researcher should try to understand the products, customers, marketing strategy, corporate strategy, company background, mission and vision of the company, etc.)
- the background of the management problem
- the nature, symptoms, and suspected causes of the problem
- the concerned stakeholders
- any assumptions about existing conditions
- possible solutions to the problem
- anticipated results of tentative solutions
- expectations from the marketing research
- any other relevant information

During the discussions, the researcher should try to find out whether the information in hand is adequate enough to specify the research problem and objectives. If not, he or she should pose more questions to the manager(s) or decision maker(s). The researcher may also approach relevant person(s) from the company to gather more information if needed. At times, interviews with industry experts (internal or external) also help in gathering critical information in understanding the management problem and defining the research problem.

Meetings and discussions can take on any of several forms, from creative and open-ended brainstorming sessions to structured workshops employing specific meeting strategies and models. Process workshops, for example, provide a forum for all members in a specific process to arrive at a process-related solution such as adding a new product to production, or implementing new protocol. War Room Sessions offer a creative space for generating ideas, diagraming, problem-solving, and testing of hypotheses. It is important when establishing meetings for the purposes of defining a research problem that all affected areas are represented in the meeting. A process workshop, for example, should bring together individuals representing all areas of the business that might be affected by the implementation of the new process, including senior management as well as the various departments and functional areas impacted.

Example of Meetings and Discussions:

- Process Workshops—Concrete Solutions designs and installs custom driveways. The company is focused on the private home market but wishes to expand into commercial contracting. As it is new to this market it decides to facilitate a process workshop to formalize its marketing methodology for this new customer base. The management problem it wishes to address is, "Can we profitably enter the commercial contracting market?" To define a related research problem it sets up a process workshop. The company brings marketing managers, production engineers, and a commercial construction consultant together to review its current data gathering procedures to determine if there are existing processes that can be modified to fit the new need. By diagramming its current processes and mapping them against the new desired criteria, the company can determine the most efficient way to expand its methods to include the new target market.
- War Room Sessions—Concrete Solutions may have a need to quickly determine the research problems associated with its management problem, "Can we profitably enter the commercial contracting market?" Because time is a critical factor, it decides to set up a war-room meeting with appropriate subject matter experts. This meeting is to take place at its head office, and will be held over the course of one day. By co-locating these experts and establishing a time limit on discussion and solution, the team is able to collectively agree that the primary research problems are, "What is the average price per square meter developers currently pay for paving?" and "What are the anticipated construction levels over the next two years?".

2.1.2.5 Situation Analysis*

One of the important steps in identifying the research problem is to conduct a situation analysis. A situation analysis involves examining the external environmental factors and internal organizational capabilities that impact how a company operates. Since companies operate in dynamic environments, understanding the changing landscape and current trends that are impacting the business helps marketing researchers identify the existing and potential problems. 5C Analysis, SWOT Analysis, and Porter's Five Forces Industry Analysis are some of the techniques used to conduct situation analyses. These analyses help marketing researchers understand their company, their customers, and the external environment in which they compete, which in turn allows them to focus on the most critical problems facing their organization.

5C Analysis

5C Analysis is one of the most popular and useful frameworks in understanding internal and external environments. It is an extension of the 3C Analysis that originally included Company, Customers, and Competitors. Collaborators and Climate were later added to the analysis to make it comprehensive. This integrated analysis covers the most important areas of marketing, and the insights generated can help identify the key problems and challenges facing the organization. However, it should be noted that not all five elements need to be considered when identifying the problem in a particular area of marketing.

Depending on the area of marketing under scrutiny, some areas need to be given more importance than others.

- **Company**—The company analysis studies an organization's vision, strategies, capabilities, product line, technology, culture, and objectives. It is useful in understanding the existing and potential problems with the company's business model and the challenges it faces from the external environment.

- **Competitors**—Competitor analysis is critical in understanding the external environment in which the firm operates. This analysis involves knowing the competitors' strengths, weaknesses, positioning, market share, and upcoming initiatives.

- **Collaborators**—Collaborators are the external stakeholders who team up with the organization in a mutually beneficial partnership. Agencies, suppliers, distributors, and business partners are typical collaborators. It is important to understand their capabilities, performances, and issues to better identify business problems.

- **Climate**—Climate analysis is the evaluation of the macro-environmental factors affecting the business. PESTEL analysis can be used to analyze climate—political, economic, social/cultural, technological, environmental, and legal scenarios are included in PESTEL.

- **Customers**—Understanding customers is a key part of situation analysis. It involves knowing the target audience, their behavior, market size, market growth, buying patterns, average purchase size, frequency of purchase, and preferred retail channels.

Example of 5C Analysis:

- A start-up shoe manufacturing company conducted a 5C analysis to identify market opportunities and work on its product pricing strategy.

  Company: While considering the base pricing of the product, the researchers considered differential pricing across the product lines, automated technologies that might reduce the cost relative to inefficient labor-intensive technologies, and the goal of the company to make cost-efficient yet trendy shoes.

  Customers: While considering the customers, the researchers took into account the market size, product benefits, market growth potential, market trends, customer buying behavior, and the image of the brand that the products are trying to create.

  Competitors: A major factor that determined product pricing was the presence and capabilities of competitors. The researchers considered the actual and potential competition, the positioning of the company's products versus that of the competition, and the strengths and weaknesses of competitors' products.

  Collaborators: The research team put special emphasis on the distributers and suppliers to understand the availability of multiple suppliers that enabled a seamless flow of raw materials at lower cost and an efficient distribution network that led to reduced prices owing to increased sales and customers acquiring better product knowledge.

  Climate: The researchers considered the macro-environmental factors, such as the political and regulatory environment, including government policies and laws regulating cartelization, taxes, and import duties. They also closely studied the economic environment including the rate of inflation, interest rates, and seasonal factors that affected product demand in a significant way.

  This elaborate 5C analysis conducted by the research team helped the researchers understand the market and customers better and helped them define the research problem which would lead to the formulation of the optimal pricing strategy.

SWOT Analysis

SWOT analysis is an important method of conducting situation analysis. The internal capabilities are studied by identifying strengths and weaknesses while the external environment is studied by identifying opportunities and threats.

The strengths and weaknesses of a company determine its internal capabilities to compete in a market and to fulfill customer expectations. A review of a company's ability to leverage its strengths and counter its weaknesses can be a good source of problem identification. Those factors that can be exploited to help the company achieve its objectives are considered opportunities, while those that hinder the company's efforts are considered threats. Marketing research is a useful tool for identifying or ratifying opportunities and for negating or preempting potential threats. For a detailed discussion on SWOT analysis, refer to SMstudy® Guide-Book One, Marketing Strategy (sections 2.1 and 2.2).

Example of SWOT Analysis:

- A leading aircraft manufacturing company, despite being the current market leader, performed a SWOT analysis to identify new opportunities, to address key threats, and to overcome weaknesses to establish itself as a long-term market leader. A recent SWOT report indicated that the company's strengths include the capability to produce commercial planes, integrated defense systems, military and missile systems, and space and communication systems. However, two weaknesses—excessive spending on R&D and a decline in the performance of integrated defense systems—had reduced the company's net earnings.

  Addressing these two issues provided the company with an opportunity to increase the production in the defense and military sectors and fulfill the market demand. Changes in government policies and contracts continue to be a threat to the company as it has to reevaluate its production and quality requirements time and again. Thus, the SWOT analysis helped the company understand that the research problem should be aimed at creating alternatives for tackling government policy changes and not only be focused on understanding new opportunities.

Porter's Five Forces Industry Analysis[3]

Porter's Five Forces model is used to analyze the long-term attractiveness of an industry. It also helps a company decide whether or not to enter an industry. The five forces include the following:

1. Threat of new entrants
2. Threat of substitutes
3. Bargaining power of customers
4. Bargaining power of suppliers
5. Competitive rivalry

[3] Porter, M.E. (1998). "The Five Competitive Forces that Shape Strategy." Competitive Advantage. New York: The Free Press.

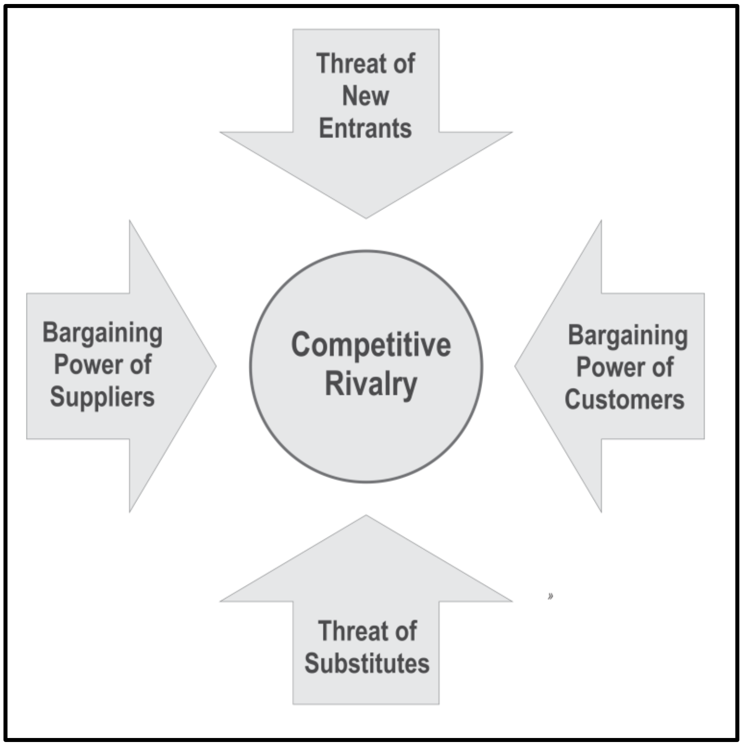


Figure 2-10: Porter's Five Forces Model

This model uses industrial organization economics to derive the five forces that affect the overall profitability of an industry. The output of the analysis can result in an early indication of the potential problems posed by the external environment. Marketing research can then be used effectively to explore and preempt these potential problems. A detailed description of Porter's Five Forces Model is provided in SMstudy® Guide-Book One, Marketing Strategy (section 2.2).

Example of Porter's Five Forces Industry Analysis:

- Porter's Five Forces model is one way to answer the first basic question in strategic management: "Why are some industries more attractive than others?" This model helps identify the key areas of the industry—the threat of new entrants, the threat of substitutes, the bargaining power of customers, the bargaining power of suppliers, and competitive rivalry. Using Porter's Five forces Model to analyze the airline industry in the United States would provide an airline with a comprehensive understanding of the attractiveness of this industry.

Threat of new entrants

The threat of a new entrant is a key component of the five forces. For the airline industry, the threat of new entrants is very low, as this industry requires an enormous amount of capital. Without a strong customer base, the new entrant may not be able to sustain itself as there might be little or no profit in the initial years. Most consumers prefer to stick with an established provider that has a proven track record in safety and quality service. Several other factors facing new entrants, such as licensing, flying experience, aircraft mechanics, and frequent changes in regulations by governmental and aviation bodies, make it difficult for new entrants. The time and money necessary to start an airline company alone prevent most people from entering the industry.

Threat of substitutes

There are substitutes to the airline industry—that is, alternate forms of transportation—including rail, road, and boat. Many people in the United States use private vehicles for long journeys. However, flying is the quickest choice for consumers, and substitutes in terms of the train and bus are minimal in their impact when expedited travel is required.

Bargaining power of customers

The customers in this industry can be classified into two groups—the individual buyer and the institutional intermediaries (i.e., travel agencies and online portals). An individual can buy a flight ticket either directly from a specific airline or indirectly through the second group of buyers. Travel agencies work with multiple airline providers to offer their customers attractive deals. The entry of low cost carriers and the resultant price wars have greatly benefited the consumer. Although there are several options available to a buyer, the purchase decisions are based on individual requirements and preferences; where some are inclined towards cost, others prefer the amenities. Overall, the bargaining power of buyers is a moderate to high threat in this industry.

Example of Porter's Five Forces Industry Analysis (continued):

Bargaining power of suppliers
The power of suppliers in the airline industry is moderate to high. An airline company depends on fuel and aircraft, which are vulnerable to the external environment. The price of aviation fuel is subject to the fluctuations of oil in the global market and is vulnerable to geopolitical and other factors. Airline companies obtain the aircraft either on an outright sale or wet lease basis from the manufacturers. Since the power is concentrated in only two major manufacturers, the supplier power is high. However, airline manufacturing itself is a high-capital driven industry and the manufacturers cannot afford to lose business over minor negotiating disputes. The airlines are the only source of income for these manufacturers, so their business is extremely important. Based on these factors, the bargaining power of suppliers has a moderate to high effect.

Competitive rivalry
There is intense competition in the airline industry. Although the threat of new entrants is minimal, the existing providers are very competitive, and most of the airlines operate in multiple geographies. The combined pressure to provide more competitive, lower-cost tickets, along with the high operational costs and other licensing costs involved, results in low profitability for many airlines. However, there are a few profit-making carriers, which are backed by large firms and governments. Compared to the other forces, competition from rivals is a major threat in the airline industry.

The output of this analysis can result in an early indication of the potential problems posed by the external environment to the airline industry. These identified potential problems can help the researcher better define the research problem.

2.1.2.6 Symptomatic Situation Analysis[4]

Symptomatic situation analysis can be used as a step in identifying a research problem. The three symptomatic situations are overt difficulties, latent difficulties, and unnoticed opportunities.

- **Overt difficulties**—These are the challenges faced by the organization that are evident and need to be addressed immediately. For example, the reduced number of downloads for an e-commerce company's app is an overt difficulty. Such difficulties present the most urgent cases for marketing research.
- **Latent difficulties**—These are the not-so-obvious challenges faced by the organization. These become evident in due time if not detected and addressed promptly. Such difficulties can be identified and resolved by marketing research, especially exploratory research.

[4] Beri, G.C, (2008). Marketing Research, 4th Ed. McGraw Hill.

- **Unnoticed opportunities**—Many opportunities are not very evident and some effort is required to identify them. Marketing research is invaluable in identifying and leveraging such opportunities.

Example of Symptomatic Situation Analysis:

- A global pizza restaurant chain provides free home delivery to customers. The traditional business model dictated that customers make a phone call to place orders. With increasing numbers of calls and other operational issues, the company developed a self-serve website and a smartphone app to make the ordering process easy and to enhance the customer buying experience.

Overt difficulties
Downtime in the mobile app or website will affect the direct sales and also result in a bad experience for the customers. These are evident challenges, and it is essential for the company to make sure the website and the app have a nearly 100 percent uptime to enhance consumer satisfaction and sales.

Latent difficulties
When customers ordered by telephone, there was an opportunity for the pizza chain to repeat the order back to the customer and confirm that it was correct. As the more automated system was implemented, the restaurant chain noticed an increase in incorrect orders. Even though they were typically due to customer error using the app, this still resulted in complaints and dissatisfaction. The pizza chain will need to research the underlying causes of these incorrect orders to seek solutions.

Unnoticed opportunities
After customers started using the app, the pizza chain noticed an increase in repeat business per customer. The convenience of ordering with a few quick clicks, appeared to minimize the chain's vulnerability to competitor's special offers and discounts. The pizza chain may have the opportunity to increase the price of their pizzas, without losing repeat business. Market research will help to determine the levels of price sensitivity, and the impact of special offers on their sales volumes.

Thus, a symptomatic situation analysis helped the researcher identify the research problems within the company that could be addressed for business stability, consumer satisfaction, and growth.

2.1.3 Outputs

2.1.3.1 Research Problem and Objectives*

Defining the marketing research problem along with its objectives accurately and completely is critical for the marketing research team to initiate the marketing research process. Improper problem definition will lead to incorrect research design selection, inappropriate data collection, and inaccurate data analysis. A detailed statement of the marketing research problem and objectives needs to be documented so that it can be used as an input for other processes within marketing research. The document can also contain information about the possible causes of the problem.

It is important to document the marketing research problem and the objectives because the research will be completely driven towards answering the research problem. The objectives must be focused and specific. The research objectives should answer the main questions such as “What is the purpose of the research?” and “What information is being sought?” and “How will the information be used?” These questions will eventually translate the management problem to the research problem statement. For example, the management problem may be whether to launch a new product. The research problem will analyze if there is any scope for the new product in a particular market. Therefore, most researchers outline the research problem, objectives, and possible research results before the research commences.

Example of Research Problem and Objectives:

- A high-end patio furniture design company is adding two new grades of furniture to its line. It may decide the primary research problem is “What are the demographics for the new low- and medium-grades of furniture to be offered?” Its objective will be to qualify and quantify these two new consumer groups in order to set appropriate price points and marketing strategies.
- An air conditioning company has had a website for many years. Its purpose is to describe the features and benefits of their product line. It sells indirectly through retail distributors and has no plans to change this, but they are interested in determining if a social media presence will increase the reputation of their brand. The company’s research problem is designed to answer three key questions: “Which social media channels are currently being used by our distributors?” “Does our target market use these channels?” “And, how much traffic do they generate?” If the research determines that there is a large enough audience seeking information on these channels, it will allocate a portion of their marketing budget to this area.
- A well-established luggage designer wants to re-investigate its value proposition. It may state the primary research problem as follows: “How is our product line differentiated from the competition on customer service, price, warranty, and quality?” Its objective will be to compare customer evaluations of the business using these criteria with evaluations of its three closest competitors using the same criteria.

2.2 Choose Research Design

Once the marketing research objectives are defined by the researcher and marketing decision maker, an approach to the problem is established and an appropriate research design is developed. The components of the research design consist of an objective/theoretical framework, analytical models, research questions, and hypotheses.

After the researcher has formulated the research problem, he or she must develop or choose a proper research design and establish the research approach. The chosen research design is a set of guidelines or a blueprint that specifies the methods and procedures for obtaining and analyzing the required information. The research design will provide a framework or plan of action for the research project. This can be called the “Master Plan” or “Project Plan” for the entire research project.

The research problem and objectives determined in the previous process need to be included as part of the design to ensure that the information obtained is appropriate for solving the problem under review. The research design will also include the sources of information, assumptions, design tools and techniques, the sampling methodology, the analytical methods with which the inputs will be treated or calculated, the budget, and the schedule of the research.

There are many alternative research designs that can satisfy research objectives. The key is to create a design that enhances the value of the information obtained, while reducing the cost and time of obtaining it. Choosing the correct research design is one of the most important steps in the research process. An incorrect research design may lead to wastage of time and money, and it may also lead to incorrect conclusions, which can have more severe implications. Figure 2-11 shows the inputs, tools, and outputs for the *Choose Research Design* process.

1. Research Problem and Objectives* 2. Senior Management Direction and Insights* 3. Expert Judgment 4. Background Information* 5. Available Information	1. Meetings and Discussions* 2. Available Information Evaluation 3. Exploratory Research Design* 4. Descriptive Research Design* 5. Causal Research Design*	1. Selected Research Design*

Figure 2-11: Choose Research Design—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

2.2.1 Inputs

2.2.1.1 Research Problem and Objectives*

While it is difficult to point to any particular process as the most important in a research project, a strong case can be made for the first step, defining the research problem and objectives. If this step is approached incorrectly, the entire research design can be misguided and unfocused.

Different types of objectives lead to different types of research designs. Selecting appropriate research design depends, to a large extent, on the research problem and objectives and how much information is already known about the problem.

The research problem may be very clear in the sense that there are strong established theories of what should be measured and how to conduct the measurements. Conversely, the research problem may lack a theoretical foundation, with the researcher trying to cope with a broad set of issues that have not been sufficiently researched beforehand and the inability to rely on existing theories. Poorly defined problems may cause confusion and hinder the ability of the researcher to develop a good research design.

Exploratory research may be appropriate when very little is known about the problem or if the objectives are not well defined. If basic information about the problem is available and the scope of the problem is limited, descriptive research is preferred. Causal research is used when the exact nature of the problem is well defined and when a relationship needs to be established between variables of interest. The research methods and tools associated with these designs will be discussed in detail in later sections.

2.2.1.2 Senior Management Direction and Insights*

When developing the research design, the researcher should take into account the views and perspectives of marketing decision makers. Senior management should have confidence that the research design will accomplish what is desired. The client’s assessment of the initial research design may warrant revision to meet the information needs of the marketing decision makers.

From their education and industry experience, senior management members may have more confidence in certain techniques, believing them to be the most effective. The researcher may need to take the perspective of the intended respondents to evaluate whether the proposed technique actually is the best means to measure or understand the issue under investigation. Thus, research design involves the researcher developing an understanding of both the type of data decision makers and senior management have confidence in and how respondents may react to certain research approaches or designs.

Every research design has some assumptions and limitations. When creating the research design, the researcher should discuss the key limitations with senior management and the marketing decision makers to convey and understand the points of view of all parties involved in establishing the research design. This information is required to avoid selecting an inappropriate course of action.

The researcher needs to understand the time and budget that management has allocated to the research project. With knowledge of time and cost constraints, the researcher can develop a research design that can be implemented within the given time and budget. At times, he or she may need to convince senior management to allocate more budget and time to implement the best possible research design. The researcher needs to ensure that constraints do not diminish the value of the research to the decision makers, or compromise the integrity of the research process. In instances where the resources are too limited to facilitate a project of sufficient quality, the firm should be advised not to undertake formal marketing research. Thus, it becomes necessary to identify resources and constraints, a task that can be better understood when discussed with the decision makers.

The researcher should understand the various legal aspects associated with the research project. While formulating the design, the researcher should confer with senior management to understand relevant legal issues associated with data collection, supplier-customer contractual relationships, privacy rights, data protection, and any possible violation of other binding obligations. For example, a particular research design may call for interviewing customers, but the organization may have a binding obligation promising not to share customer data with third parties, such as the researcher. In such a case, the researcher may need to look at altering the research design. The researcher should never put the organization in legal jeopardy while conducting its research. Senior management or the legal department of an organization should review the procedures to make sure no legal problems are created by implementing the proposed design.

Also, an organization may not want the researcher to reveal its identity while the research is being conducted. The identity of the survey sponsor and/or the ultimate client for whom a survey is being conducted is typically held in confidence at all times, unless this identity is to be revealed as part of the research design. The researcher should also understand from senior management the confidentiality of the research information collected during the research project.

The researcher should discuss senior management's expectations regarding delivery of interim reports, who in the organization may be helpful for supplying further information, and which source materials and individuals are needed to successfully complete the research. While choosing the research design, the researcher should seek the inputs of the sponsoring party as and when required.

Examples of Senior Management Direction and Insights:

- A leading online marketplace had been conducting consumer research on buyers from its website, but had not been focusing on sellers. This omission was highlighted by members of senior management when discussing the research design. Subsequently, the company restructured the research design to address questions related to needs of sellers, resulting in more product availability on the website. The sellers are also consumers of the website, and thus, they cannot be ignored while conducting consumer research. The senior management's perspective provided critical input identifying the sellers as a second consumer segment which proved to be very important in the research design.
- A seismic measurement company planned to conduct marketing research to study the potential operators involved in deep-well drilling in a specific region. As part of the secondary data collection of the research design, the marketing team had planned to review publically available drilling data released by the government to identify which companies are involved in the area and what depths or strata they are targeting. Senior management was approached with this suggestion. Based on their knowledge that operating companies are not legally required to release certain information for the first five years after starting a drilling program, the marketing team realized that the data obtained in this fashion might not be current or valuable.

2.2.1.3 Expert Judgment

It is a well-established fact that there is no single best research design for any particular research project. The researcher often has several alternative research designs that can accomplish the stated research objectives. For example, a consumer durables company wants to predict its market share for the next few years by conducting a research project. In this case, the researcher has several options to consider: surveying dealers and distributors, collecting industry-wide sales figures from the last few years and analyzing the collected data, surveying a large sample size of future potential customers, analyzing driving factors of the market, monitoring industry trends, and others. Any one of these options may yield a reliable forecast.

The ability to select the most appropriate research design develops with experience. Inexperienced researchers often choose a particular method over others because of their comfort with the method. Researchers should seek opinions from industry experts and research experts who can provide valuable inputs in choosing the best research design within the project's given constraints.

2.2.1.4 Background Information*

Background information is an integral part of the research design. Background information puts the research objectives into context, helping the researcher understand why certain research objectives are being pursued. The background information also gives a framework for the researcher to investigate other potential events and contributing factors or causes in addition to the defined research objectives. The background information provides hints to the researcher regarding what information he or she should be looking for and where to look for it. Detailed background information regarding the problem under study also helps the researcher to adopt a particular research design over others.

Example of Background Information:

- A company markets a labor-intensive product distributed in multiple global markets that is very price sensitive. It has been noticed that sales are steadily declining in one of the market regions and the company's objective is to identify the problem and take countermeasures to revive the market in this region. The strength of the sales and marketing has not changed over the past year; there have been no changes in government regulations; and the cost of raw materials, labor, fuel, and other incidental costs have not changed significantly in this region. Some two-year-old data are available regarding consumer demographics and preferences; this data was used during the marketing strategy formulation the previous year.

  Based on this background information, the researcher realized that the correct research design will depend on multiple parameters, such as changes in industry trends, changes in customer preferences, changes in competitive pricing, and the entry of new competition, and a good starting point for this project may be studying secondary research to gather more data on these parameters.

2.2.1.5 Available Information

In the selection of design, researchers must take into account the availability of internal and external resources. There may be data available in the form of published material, online databases, or information provided by syndicated services. The collection and analysis of secondary data helps define the problem of market research and develop a strategy. Examination of available secondary data is a prerequisite to the collection of primary data. The researcher should consider collecting primary data only when the secondary data sources have been exhausted or yield partial information. Therefore, secondary data are key components of successful research design. Secondary data can help in sample design and in finalizing the details of primary research methods. Given the explosion of available secondary data sources, it is possible for a researcher to access enough data to solve a specific market research problem.

The researcher should also look for reports based on previous similar research projects (in-house or external). In any research project, lessons learned from previous similar projects play an important role in deciding appropriate research methods.

Example of Available Information:

- If a company is undertaking market research with the objective of identifying the best pricing strategy for its product, the analysis of existing data on the competitors' pricing by geographic area can be a good starting point. Instead of initially focusing on primary research, the researcher can access information such as competitor pricing details, which can be obtained from published product catalogues. These inputs can help the researcher construct the primary research design to focus on more hands-on data such as current consumer preferences and buying behavior.

2.2.2 Tools

2.2.2.1 Meetings and Discussions*

It is essential for a researcher to discuss the research project with stakeholders to ensure the findings are helpful. These meetings can be conducted within the organization or by seeking insights from experts externally. These discussions are conducted at the early stages of the research project and at frequent intervals to evaluate the research process.

The researcher brings a small group of stakeholders and experts to discuss the key research problem and to gain information relevant to it. The outcome of these meetings can help the researcher define the research process and priorities. The researcher gathers key information from these discussions, which enables the research team to choose a suitable research design—exploratory, descriptive, or causal. These initial meetings and discussions help the researcher clearly define the research problem. The key factor of these meetings is to ensure that the goals are met using the appropriate research design.

Example of Meetings and Discussions:

- As described in section 2.2.1.4, if a company selling a labor-intensive product experiences declining sales in a certain geography, the researcher may start by setting up discussions with the sales and marketing teams across the geographies. The collective inputs from the internal sales and marketing teams will provide significant information on comparative sales scenarios with other markets, historical data analysis of fluctuating buyer behavior in this industry, and potential causes for the decline in sales. The researcher can also plan meetings with industry experts to understand what has changed in the recent past that can bring about changes in the buying behavior of customers. This may validate the learning from the sales and marketing teams. The information collected from these meetings can be enough to provide the researcher with a good understanding of the basic problem and its causes. He or she can then focus on an exploratory research design to examine the causes of the declining sales.

2.2.2.2 Available Information Evaluation

Every research problem is unique in its own way, but almost all research problems and objectives can be matched to one of three types of research designs—exploratory, descriptive, or causal. Available information evaluation is one of the important tools for choosing the research design. Available information such as the nature of the problem, scope of the problem, objectives, and known information, informs the decision maker regarding which type of research design to adopt. Table 2-1 shows the three types of research design and how their selection depends on the available information.

Available Information	Research Design
Exploratory research design is selected when the exact nature of the problem is not known and limited information is available about the problem. Its goal is to gather information, to gain clarity, to discover ideas, and to suggest possible solutions.	Exploratory
Descriptive research design is used when the basic information about the problem is available and the scope of the research is limited to a few aspects that need to be examined.	Descriptive
Causal research design is used when the exact nature of the problem is known and a relationship needs to be established between the variables of interest. It is usually associated with limited scope and high information availability.	Causal

Table 2-1: Types of Research Design

2.2.2.3 Exploratory Research Design*

As previously discussed, research design is classified into three major categories—exploratory, descriptive, and causal. The researcher's choice of design depends on the objectives of the research or the research problem. Exploratory research design is chosen to gain background information and to define the terms of the research problem. This is used to clarify research problems and hypotheses and to establish research priorities. A hypothesis is a statement based on limited evidence which can be proved or disproved and leads to further investigation. It helps organizations to formulate their problems clearly.

Exploratory research design is conducted for a research problem when the researcher has no past data or only a few studies for reference. Sometimes this research is informal and unstructured. It serves as a tool for initial research that provides a hypothetical or theoretical idea of the research problem. It will not offer concrete solutions for the research problem. This research is conducted in order to determine the nature of the problem and helps the researcher to develop a better understanding of the problem. Exploratory research is flexible and provides the initial groundwork for future research. Exploratory research requires the researcher to investigate different sources such as published secondary data, data from other surveys, observation of research items, and opinions about a company, product, or service.

Example of Exploratory Research Design:

- Ideafire is a one-year-old e-commerce startup company in the education domain. The company sells customized self-learning courses to corporations and universities. Since its inception, the company has been exceeding expectations by achieving a high sales growth rate. However, after the first ten months, sales suddenly started dropping. Due to the lack of any past data or history, the sales director was puzzled about the reasons for this decline in sales. Rather than making assumptions, he appointed a marketing research agency to conduct an exploratory research study in order to discern the possible reasons. The objective of this research was not to figure out a solution to the declining sales problem, but rather to identify the possible reasons, such as seasonality, competition, or ineffective marketing, and to better understand the factors affecting sales. Once these potential causes are identified, the strength of each reason can be tested using causal research.

Uses of Exploratory Research Design

Some uses of exploratory research design are as follows:

- **To gain background information on the research problem**—Background information is essential when the researcher is left with little or no information about the research problem. In such instances, the researcher will look for Management Information Systems (MIS) data or past data that can provide useful insights. The background of the company, brand perception, quality, and sales figures provide information that the researcher may find useful in formulating the research process.

- **To define terms**—Exploratory research design helps to define terms and concepts of the research problem. For example, the brand image of a service or a product needs to be defined by several components such as price, product range, quality perception, customer feedback, and after-sales service. The researcher can use the exploratory research design to categorize these components and address the branding issue of the product or service.

- **To clarify research problem and hypothesis**—It is essential for a researcher to identify or define the research problem accurately, and this enables him or her to arrive at a hypothesis for future study.

- **To establish research priorities**—A research project can involve several research studies such as case studies, in-depth interviews, focus groups, and customer feedback. The researcher can prioritize the importance of these studies using exploratory research design.

Methods of Exploratory Research Design

The common methods of exploratory research design are as follows:

- **Secondary Research Techniques**
  - **Secondary Data Analysis**—The researcher may find some existing information relevant to the research problem from sources such as journals, books, MIS, and the Internet. The process of searching and interpreting the existing information is known as secondary data analysis. An analysis of the secondary data is often referred as the core of exploratory research.

Example of Secondary Data Analysis:

- A leading telecom service provider needs to decide the next pack of value-added services to offer to its subscribers. Therefore, the company decides to conduct an exploratory secondary research study to understand the main services currently used by consumers and the need gaps in the market. The consumer needs that are not currently being addressed by other players, such as pay-per-use models and integrated service offerings, would be the provider's main focus area when developing new value-added services combinations.

  - **Case Studies**—Case studies are an effective source of information to aid a researcher in addressing a research problem. The researcher can refer to available information in a past case study that is relevant to the present research problem.

Examples of Case Studies:

- Manufacturing companies often require exploratory research to improve their operations processes. These types of research projects are mostly secondary, where they examine case studies of similar companies that have optimized their operations process in similar situations.
- Medical practitioners also refer to case studies when they face some difficulty that they have not faced before. Past case studies can show them a path taken to treat a patient successfully, which can be repeated with a patient showing similar symptoms.
- Technological advancements often provide new opportunities. By reviewing case studies of how a new technology has succeeded in a parallel industry, a company can gather pertinent information to design a research strategy for its own application. If a company wishes to launch a pay-per-view service for instructional videos, it may review successful and failed attempts of similar companies in the entertainment industry. By understanding the user's acceptance criteria of this model in a related field, the instructional video provider can focus on the relevant research criteria.

- **Observation Techniques**
  - **Structured**—Structured observation, also referred to as systematic observation, is an observational data collection technique in which the researcher collects the information directly without the mediation of respondents, interviewees, and so on. In this structured technique, the data is collected according to predefined rules.

Examples of Structured Observation:

- A grocery chain may wish to understand the factors influencing impulse purchases. It can set up a structured observation experiment to record sales levels of purchases based on product placement in the stores. For the first month, the company may place a display of snack foods at the store's entrance, and then record the total volume of purchases. The second month, the company may place the same snack foods, in the same display, next to the cash registers. By comparing purchase volumes, it can observe the impact of product placement on impulse buying.
- A kitchen appliances company wants to understand the need gaps in the cookware category among its target consumers. To achieve this objective, a research study was designed where consumers were asked to prepare food items using the cookware under study. The issues faced and the challenges experienced, sometimes unnoticed by the participants, were observed by the research team, and a product extension concept was formulated based on this research.

  - **Unstructured**—In unstructured observation, the researcher enters the field with some general ideas of what might be salient, but not of what specifically will be observed. Therefore, observation is holistic, unstructured, and unfocused, with the investigator attempting to document as much as possible about the setting and its participants in order to discover themes of interest.

Example of Unstructured Observations:

- Often, consumer durables companies seeking to discover new features that can be implemented and launched in the market undertake exploratory research studies to understand areas of discomfort or challenges faced by consumers with currently available products. This research can be a qualitative research study where researchers talk to consumers directly and ask about features required or about improvements that can be made to certain products. Companies often conduct unstructured observations of consumers using their products to see the difficulties that they face or areas of effort that can be avoided or can be managed more efficiently.

  - **Ethnographic**—In ethnographic research, a researcher observes or interacts with a particular geography or ethnicity to determine consumer behavior. Understanding consumer rituals is essential in determining the marketing plans for a particular geography. Ethnographic research can help companies gain popularity and increase sales by understanding consumer rituals and preferences.

Examples of Ethnographies:

- Multi-national food chains like McDonalds, KFC, and Subway have varied menus to cater to different geographies or ethnic groups. By observing and interacting with consumers and testing various flavors and options, such chains are able to identify products that appeal to consumers in certain geographic regions and tap into unique and specific consumer preferences in order to gain popularity in these markets.
- Regional holidays and celebrations are commonly used to increase sales of specific types of fast moving goods. A chocolate bar manufacturer will be aware that candies packaged with an Easter theme will have a high acceptance rate in geographies that celebrate that holiday, yet will have limited impact in regions that do not. By reviewing census reports and other demographic data, the manufacturer can focus on areas with high concentrations of its target audience.

- **Qualitative Techniques**
    - **Focus groups**—Groups recruited to discuss the current research problem can produce insightful information and possible suggestions to support the marketing research process.

Example of Focus Group Discussion:

- A credit card company wants to understand the expectations of consumers in the market. It organizes focus group discussions among its consumers, and those of its competitors, to understand what the competitor is offering and what can be offered by the company to make its customers happy and to attract new customers. Its offers, benefits, and reward points system need to be tailored to the most used and wanted services among the target consumers; these can be anything from benefits given on payments of utility bills to offers on dining, movie tickets, and more.

    - **In-depth Interviews**—These interviews are often unstructured and conducted by the researcher to gain expert insights or to gain consumer opinions through door-to-door interviewing. Although these interviews are unstructured, the questions and probing are focused on the research problem. The main focus of in-depth interviews is to seek answers and possible solutions for the research problem.

Example of In-depth Interviews:

- A new fashion retail portal for women wants to understand the aspirations of its potential customers—women who currently buy from other fashion portals. The potential customers' needs and aspirations are best disclosed in one-on-one discussions with company representatives who are responsible for deciding the product range for the portal. The knowledge gained from the interviews can also help the marketing team plan communications that will appeal to the women in their target group and connect with them on an emotional level. In-depth interviews prove to be a powerful tool in this type of scenario.

- **Projective Techniques**—Projective techniques may be classified as a structured, indirect way of investigating the "whys" of a situation.[5] Projective techniques are not used to measure, but to understand the attitudes, beliefs, and motivations of the consumer. Projective techniques help the researcher to understand customer perception of a product or a service.

Example of Projective Techniques:

- Projective techniques can be used to understand the desired personal values consumers seek in the products they are evaluating. A focus group might be asked questions such as, "What type of person would drive this car?" Framing questions in this fashion gives the observer information related to the qualities that a consumer projects upon a product or an environment.

- **Quantitative Techniques**

  - **Exploratory Multivariate Analysis—**This involves analyzing more than two variables or factors. The most favorable scenario in this case is one dependent variable and multiple independent variables; that is, one particular factor is influenced by many other factors. Another scenario could be two or more variables that are dependent on other independent variables.

Examples of Exploratory Multivariate Analysis:

- A multivariate analysis can be conducted to determine the impact of volume discounting on purchasing decisions. A test group can be given pricing options for the same related products. The group will be informed of the specific price for each item. They will then be given the option of purchasing based on a volume discount per item, or a discounted, bundled price for all items. This will help to quantify the impact of each pricing strategy and clarify the customer's perception of the individual and bundled value of the items.
- Most fast moving consumer goods companies want to understand the impact of season, temperature, holidays, promotions, and the like on the sales of their products. To understand these relationships between the dependent variable "sales" and all other independent variables, they carry out a secondary research study followed by data analysis in which data from all in-store promotions and the sales generated are used. These data are regressed along with other time series data to understand the impact of all the variables on the dependent variable, sales.

  - **Surveys**—Surveys can provide quality information relevant to the research problem. The surveys can be conducted, usually within a target group of past customers or people who have enough knowledge about the product or service. In addition to standard surveys, there are two specific surveys that have unique target groups. These are as follows:

---

[5] Webb, John. R. (2002). Understanding & Designing Marketing Research. London, Thomson Learning.

- **Pilot survey**: In a pilot survey, data are collected from a smaller sample group relative to the planned sample group. A pilot survey is a good option to test a survey before expense and time is invested in a larger more comprehensive survey.

Example of Pilot Survey:

- A multi-national photocopier manufacturer is developing a low cost 3-D printer. The marketing department is researching potential industrial applications for this technology. The technology is new to the market and the potential consumer group is broad and undefined. It decides to use a pilot survey targeting appliance repair companies. By limiting the initial survey, it can measure the response, minimize the quantity of variables, and gain an understanding of how to effectively proceed with future surveys and data collection.

- **Expert survey**: This survey involves collecting data from a group having experience with the particular research problem.

Example of Expert Survey:

- A professional association has been holding annual conferences to increase membership and build relationships within the association. It plans to hold events in multiple locations in the future. As the conferences have been well attended in their current location, the association may distribute a survey to the previous attendees to help determine the factors that have contributed to the popularity of past conferences so that these specific criteria can be met in the new locations.

2.2.2.4 Descriptive Research Design*

Descriptive Research Design is used to determine specific characteristics of a group, people, or organization associated with the marketing research problem. Descriptive research is most suitable if the researcher wants to know who the customers are, where they are from, and what they want. It provides answers to the questions who, what, when, where, and how. In this research design, basic information about the problem is available and the scope of the problem is limited to a few areas that need to be examined. If the sample is representative of a larger group, the research findings can be used to make predictions. For example, predicting sales for a geographic region, brand, and product typically involves monitoring behavior at a specific place and time. It does not necessarily answer questions related to how and why something is happening.

Example of Descriptive Research Design:

- An online education publisher wants to launch a new range of self-learning courses for the healthcare industry. In order to prepare the marketing strategy, the marketing manager needs to clearly understand the target segment and its characteristics. The manager launches a descriptive marketing research study to understand different characteristics and preferences of the target market. Based on the inputs of the marketing research, the marketing manager will be in a better position to define the positioning for the product.

Descriptive research design involves observing and monitoring the behavior of a sample group from the target population. Depending on the research objectives, the marketing researcher can choose the type of sampling strategy. Sampling design is discussed in detail in section 3.2.1.5. There are two categories of descriptive research design: cross-sectional and longitudinal designs. Figure 2-12 shows the two types of designs and the subtypes of each.

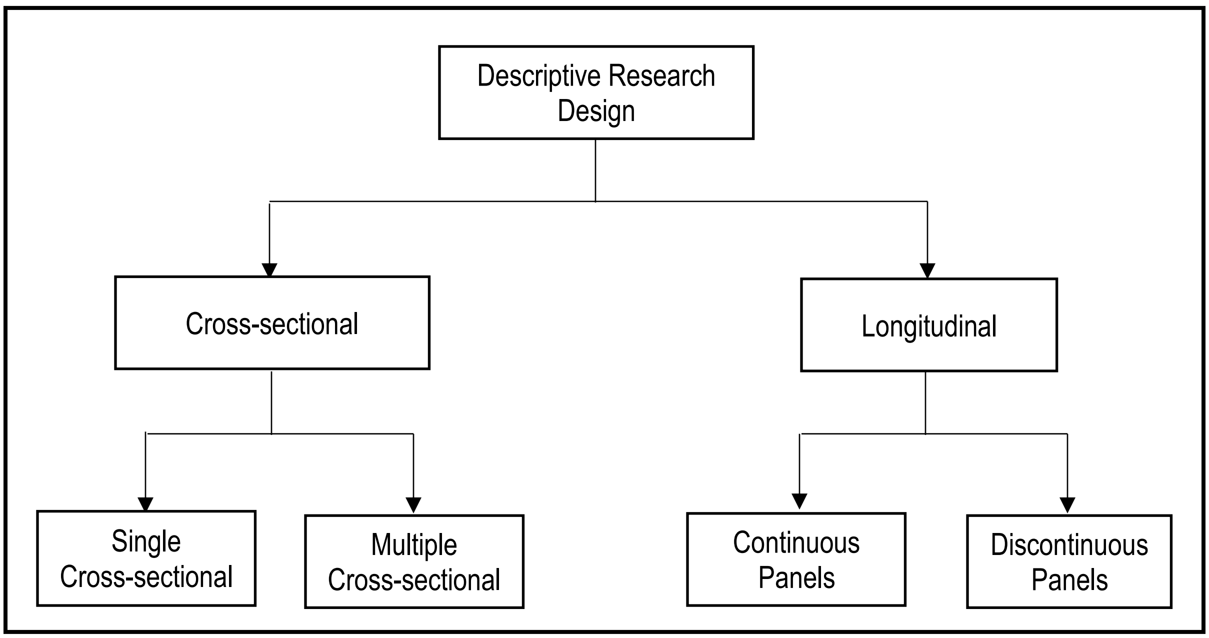


Figure 2-12: Descriptive Research Design Model

1. **Cross-sectional Design**—In cross-sectional design, the collection of information from the sample group happens at one point in time. Research to understand the customer preference for a particular brand while purchasing garments during a festive season would fall under cross-sectional design.

    - **Single Cross-sectional Design**—This approach refers to a research design in which only one sample group from a larger target group is used for collecting information. In order for the researcher to obtain accurate results, the sample group selected should be representative of the target group and share similar characteristics, such as social and educational status.

Example of Single Cross-sectional Design:

- A company wants to determine which of two variants of their product is better perceived within a given market. The company can use the single cross-sectional design approach and provide the two product variants to a sample group to determine their preference.

- **Multiple Cross-sectional Design**—When two or more groups are used from the target group to collect information, this type of research is referred to as multiple cross-sectional design. The collection of information happens only once in each sample. Sample surveys, usually conducted by various television news channels, are multiple cross-sectional designs where the sample groups are representative of the larger target group. Usually, only a single variable is examined among multiple sample groups.

Example Multiple Cross-sectional Design:

- A company wanting to determine the percentage change in consumption of a particular product with age, can use the multiple cross-sectional design approach. Sample groups can be selected from four age groups such as 20–30 years, 30–40 years, 40–50 years, and 50–60 years. The percentage consumption of the product in each group can then be calculated. The variable considered here is age and by selecting similar sample groups, the change in consumption of the product can be attributed to age.

2. **Longitudinal Design**—In longitudinal design, the same sample group is examined and the information is collected over an extended period of research. This is helpful for researchers who want to track changes in the sample group over a period of time. Tracking changes in the same group, unlike cross-sectional design, reduces the impact of many external factors on the information collected. This makes tracking changes more accurate. Since the same sample group, referred to as a "panel," participates in the research, availability of all group members throughout the process is important for accurate analysis of the research. Because this type of research is extended over a long period of time, it takes a lot of effort and can be very expensive. Due to the high costs, small sample groups are usually used for this research. The smaller sample size can make it difficult to apply the results to a larger target group, as the sample group may not always be representative of the total population. Longitudinal design is classified based on the type of panel—continuous or discontinuous[6]—used in the study:

[6] Burns, Alvin C. and Bush, Ronald F. (2009). Marketing Research – Fifth Edition.

- **Continuous Panels**—In a study that uses a continuous panel, the questions asked of the panel members remain the same over the research period. This is helpful in problems where the researcher wants to understand the changes in customer behavior over a period of time.

Example of Continuous Panels:

- A researcher can track purchase information of a grocery store customer by asking the same questions over a period of time. The results for this research can provide insights into why the panel members changed product brands over time.

- **Discontinuous Panels**—In a study that uses a discontinuous panel, panel members respond to different questions at different points in time over the research process. Since the panel members are available for the extended duration of the research, discontinuous panels act as a ready source of information for the researchers to address a wide variety of problems. In discontinuous panels, the demographics and characteristics of the panel members are representative of the target population. This ensures the accuracy of the results when applied to the larger group.

Example of Discontinuous Panels:

- Online communities managed by e-commerce players act as discontinuous panels where the e-commerce player can ask any set of questions at any time to obtain responses from its customers and make decisions about their offerings accordingly.

  These communities are often used to test new concepts where an outside opinion is required to refine a new line of products or new marketing strategy.

There are both advantages and disadvantages to using cross-sectional and longitudinal design approaches. Some of these are provided in Table 2-2.

	Cross-sectional Design	Longitudinal Design
Advantages	• Representative sample group • Unbiased response of participants	• Higher accuracy of data collected • Ability to track changes over time
Disadvantages	• Limited data: data are collected at only one point in time from one target group • External factors and variables can impact data collected	• High attrition rate of participants • More effort and time required and can be expensive

Table 2-2: Advantages and Disadvantages of Cross-sectional and Longitudinal Designs

2.2.2.5 Causal Research Design*

The third type of research design is causal research design. Causal research design is used when a cause-and-effect relationship has to be established between two variables related to the research problem. Put simply, when one independent variable (X) affects the state of another dependent variable (Y), there is a causal relationship. For example, if sales of an online retailer increase during discount periods, then the increase in sales can be attributed to the discounts. Thus, discounts and increase in sales have a causal relationship. In popular view, discounts and increase in sales are thought to imply a deterministic relationship. However, in scientific interpretation, it implies a probabilistic relationship. Any increase in sales cannot be solely attributed to the discounts. There can be a number of variables that might affect the increase in sales. Even though the effect of such variables can be negated or controlled for in an experimental setting, it can never be entirely ruled out.

To establish a causal relationship, three conditions need to be fulfilled—concomitant variation, time order of occurrence, and elimination of extraneous variables. It should be noted here that these three conditions are necessary for causality, but are not sufficient.

- **Concomitant Variation**[7]—One of the important conditions for causality is concomitant variation. It is the extent to which the two variables in a cause-effect relationship vary together in a predictable manner. Suppose an online retailer offers discounts on some product categories and not on others. At the end of the year, it is found that higher sales have occurred for categories with discounts while low sales were observed for categories with no discount. It can be said that discounts and sales have a concomitant variation.

- **Time Order of Occurrence**—Time order of occurrence of the cause and effect is of critical importance for causality. The cause X should precede or occur simultaneously with the effect Y. An increase in sales should happen after the discounts are applied, if the discounts and sales are in a causal relationship. A past history of increases in sales (an effect of historical discounts) during the discount period can motivate (cause) a company to offer future discounts. Thus, a variable can be both cause and effect in certain cases.

- **Elimination of Extraneous Factors**—To study the effect of an independent variable on the dependent variable, it is imperative that the extraneous variables are controlled. A causal relationship can only be established if *all* the other variables except the ones under study are kept the same. In order to establish the causality between discounts and sales, for example, it is required that the other variables such as standard pricing, product, branding, and seasonal variations are kept constant.

---

[7] Clow, Kenneth, E. and James, Karen E. (2014). Essentials of Marketing Research – Putting Research into Practice. Sage.

Experimental design is one of the primary methods of causal research. In experimental design, independent variables can be manipulated in a controlled environment. This ensures that the effect of other variables on the dependent variable is minimized as much as possible. The experimental design process specifies the following requirements:

- an independent variable to be manipulated in the experiment
- test units to be exposed to the treatment
- a dependent variable to be measured
- method of controlling the extraneous variables

The various components are defined as follows:

- **Dependent variable**—The dependent variable is the outcome of the experiment that is measured by the researchers. Sales, awareness, and profit are typical dependent variables.

- **Independent variable**—An independent variable is the trigger that leads to the effect. Also called *treatments*, these variables are manipulated as part of the experiment, and the effect on the dependent variable is measured. Advertising expenses, discounts, and price levels are typical independent variables.

- **Test units—**Test units are individuals, groups, or entities that are exposed to the experiment to measure their response on the dependent variable. Customers are the test units of the online retailer in the example used previously.

- **Extraneous variables—**Any variable other than the intended independent variable that can affect the behavior of the dependent variable is called an extraneous variable. Extraneous variables need to be controlled in an experiment to study the effect of the independent variable on the dependent variable.

- **Pretest**—The measurement that is taken before the exposure of test units to the independent variable is referred to as the pretest.

- **Posttest**—The measurement that is taken after the exposure of test units to the independent variable is considered the posttest.

- **Experiment**—The alteration or manipulation of an independent variable to measure the impact on a dependent variable in a controlled setting constitutes an experiment.

Example of Variables in Experiments:

- A leading retail chain in the UK tries to understand the impact of an in-store promotion on the sales of its private label snacks brand. It decides to conduct a series of experiments, in which it would run a promotion for a period of one week and assess the impact on sales of the product at three different periods in the year. It observed that although the incremental sales due to the in-store promotion varied in all three periods, the promotion had some positive impact on sales. The company then realized that the variation in incremental sales was due to the season and not due to the high or low impact of the promotion. The promotion would have the same impact on sales if the seasonality were removed.

  In this experiment, the product sales data set was the dependent variable, the promotion was the independent variable, seasonality was the extraneous variable, and the private label product was the test unit.

Experimental Design Notation

Experimental design is described and best understood with the use of standard symbols and structure to represent the various components, activities, and composition of the experiment. These symbols and standard syntax are as follows:

- X—represents the exposure of test units to the independent variable.
- O—represents the process of observation of the dependent variable.
- R—represents the random assignment of the test units to separate treatments.
- Movement from left to right indicates movement through time.
- Symbols in a single row pertain to a specific treatment group.

Symbols in a particular column refer to the events that occur simultaneously.

Types of Experimental Design

Experimental design is often classified as either true experimental design or quasi-experimental design. These are defined as follows:

- **True Experimental Design**—An experimental design that eliminates the effects of the extraneous variables is called a true experimental design. Control groups are used to achieve this effect.

- **Quasi-Experimental Design**—An experimental design that does not control the effects of the extraneous variables is called a quasi-experimental design.

In addition to the classification of true experimental and quasi-experimental designs, several subtypes of experimental designs are distinguished from one another based on several of the factors previously defined. Among these are After-only Design, Before-After Design, and Before-After with Control Group Design.

1. **After-only Design**—This is the simplest of all the experimental designs. The independent variable is exposed to the alteration and the result is observed after a period of time. It is represented as below:

   X O

   Where,
   X indicates the manipulation of the independent variable
   O represents the observation or measurement—a posttest—of the dependent variable

   The gap between the two symbols is representative of the time span between the cause and the effect. This is a quasi-experimental design as the effect of the extraneous variable is not controlled. In the example of the online retailer, X is the discount applied on a particular category of products and O is the increased sales. The drawback of this particular design is that it does not measure the impact of the change in independent variable but only measures the final state of the dependent variable.

Example of After-only Design:

- A car dealership wants to increase sales, but first needs to build awareness in its local area. To do this, the dealership manager decides to advertise in a local newspaper.

  - X indicates the published advertisement that would increase awareness of the enterprise.
  - O represents the observation or measurement—a posttest of the dependent variable awareness— achieved by conducting a door-to-door household survey.

  After-only design: X O

  The advertisement is paid for and published. The day after the ad is published, the survey is conducted to measure awareness levels.

2. **Before-After Design**—In this design, first the dependent variable is measured, then the independent variable is altered and finally the dependent variable is measured again to understand the impact of the treatment on the test units. This design differs from the after-only design in that it has both a pretest as well as a posttest. It is represented as below:

   $\mathbf{O_1 \quad X \quad O_2}$

   Where,
   X indicates the manipulation of the independent variable

O_1 represents the measurement of the dependent variable before the manipulation of the independent variable, a pretest
O_2 represents the measurement of the dependent variable after the manipulation of the independent variable, a posttest

The advantage of this design is that it measures the effectiveness of the experiment by measuring the change in the dependent variable. $O_2 – O_1$ is the change in the dependent variable and represents the outcome of the event. In the example of the online retailer, $O_2 – O_1$ is the increase in sales due to the discounts applied. This is also a quasi-experimental design as there can be multiple extraneous variables that can impact the change in the dependent variable.

Example of Before-After Design:

- A car dealership wants to increase sales and wants to first build awareness in the local area. To do this, the dealership manager decides to advertise in a local newspaper.
  - X indicates the published advertisement that would increase awareness of the enterprise
  - O represents the observation or measurement—a posttest of the dependent variable awareness— achieved by conducting a door-to-door household survey .

  Before-After Design: O1 X O2

  In the days prior to the ad being published, a survey is conducted measuring the awareness pre-advertisement (O1). The advertisement is paid for and published. The day after publication, the survey is conducted again to measure awareness levels following the advertisement.

  The awareness levels following the advertisement and how effective the advertising expenditure was can be inferred from the pre-ad and post-ad awareness levels.

3. **Before-After with Control Group**—Extraneous variables are controlled by using another test group called the “control group.” The control group is not exposed to the manipulation of the independent variable. By comparing the results of the control group and experimental group, the effects of extraneous variables can be controlled.

   As part of the before-after design with the control group, the test group is divided into two homogenous groups—a control group and an experimental group. A pretest measurement of the dependent variable is taken for both the groups. The experimental group is then exposed to the change in the independent variable. A posttest measurement is then taken for both the groups. It is represented below:

Experimental Group	O_1	**X**	O_2
Control Group	O_3		O_4

Where,
X indicates the manipulation of the independent variable
O_1 represents the measurement of the dependent variable in the experimental group before the manipulation of the independent variable in the experimental group, a pretest
O_2 represents the measurement of the dependent variable in the experimental group after the manipulation of the independent variable in the experimental group, a posttest
O_3 represents the measurement of the dependent variable in the control group before the manipulation of the independent variable in the experimental group, a pretest
O_4 represents the measurement of the dependent variable in the control group after the manipulation of the independent variable in the experimental group, a posttest

The treatment effect is measured as $(O_2 – O_1) – (O_4 – O_3)$.

The extraneous variables are controlled as follows:
$(O_2 – O_1) = TE + a + b + c + d + e + IT$
$(O_4 – O_3) = a + b + c + d + e$
$(O_2 – O_1) – (O_4 – O_3) = TE + IT$

Where,
a, b, c, d, e are the extraneous variables.
TE – Testing effect
IT – Interactive testing effect

Extraneous variables are cancelled out when the outcomes of the control group and experimental group are compared. A pretest measurement affects the response of the test group to the experiment. This is called the interactive testing effect and it cannot be controlled. This design is an improvement over the before-after design and is an effective way of controlling the extraneous variables. Taking the earlier example, suppose the online retailer divides the incoming website traffic into two comparable groups as part of the A/B testing and provides a discount only to one group. The group that received the discount is the experimental group and the other group is the control group. Thus, the testing effect is the increase in sales of the group that received the discount minus the increase in sales of the control group.

Example of Before-After with Control Group:

- A car dealership wants to increase sales and wants to first build awareness in the local area. To do this, the dealership manager decides to advertise in a local newspaper.
  - X indicates the published advertisement that would increase awareness of the enterprise.
  - O represents the observation or measurement—a posttest, of the dependent variable awareness—achieved by conducting a door-to-door household survey.

  Before-After with Control Group

Experimental Group	O1	X	O2
Control Group	O3		O4

  The same survey measuring awareness pre- and post-ad placement is undertaken in one area of the city.
  In another area where there is no circulation of the local newspaper, the same surveys are conducted at the same time as the pre- and post-surveys with the experimental group.

  Any differences between the experimental and control group in the pretest must be understood as being due to factors other than the ad.

  A typical example of Before–After with Control group is “test marketing” done only in one region before rolling out on a larger scale.

Validity of Experiments

A critical component of experiments is ensuring validity. The validity of an experiment[8] is the extent to which the experiment measures what it is supposed to measure. Experiments can be deemed valid if the following two conditions are satisfied:

- if it is established that the change in dependent variable has indeed occurred due to the independent variable and not due to other extraneous variables
- if the results of the experiment apply to the larger population of interest or to other populations of interest

The first condition is termed as internal validity and the second is referred to as external validity. These are defined as follows:

[8] Malhotra, Naresh K. (2008). *Marketing Research: An Applied Orientation*.

Internal Validity—Internal validity refers to the extent to which the manipulation of the dependent variable can be attributed to the independent variable. As discussed earlier, extraneous variables can distort the results of an experiment by affecting the dependent variables themselves. It is important that the effect of such variables is controlled in order to study causality.

Example of Internal Validity:

- In the car dealership example used in the before-after with control group design, the control group and the experimental group of survey respondents were assumed similar. But in reality, they might have different demographic, geographic, psychographic, and behavioral traits. Thus, one or more extraneous variables are introduced in the design and distort the results of the experiment. Extraneous variables need to be controlled in order for experiments to have internal validity.

External Validity—External validity refers to the extent to which the results of the experiment apply to the real world. This depends mainly on how well the test unit represents the actual population. Proper care needs to be taken while selecting the sample units. Otherwise the results will not be valid for the actual population under study. Another threat to external validity is the artificial setup of the experiments. Since experiments are designed to eliminate the effects of extraneous variables, the setup does not match the real world and, therefore, the results might not be valid in the real world. Thus, there is a tradeoff between designing the setup to ensure internal validity and designing it to ensure external validity.

Example of External Validity:

- In the car dealership example used in the before-after with control group design, the survey was conducted within a small, regional geographic area. If the survey results measuring the effectiveness of newspaper advertising on awareness of the dealership can be replicated in other regions or a larger geographic area, the study has external validity.

2.2.3 Outputs

2.2.3.1 Selected Research Design*

It is important to select a research design that will be used to undertake the marketing research based on the research problem, the objectives, the information available, and the information required. Researchers will need to assess whether it is feasible to collect the data and whether the study can be constructed with validity and reliability. Other considerations include whether it is ethical to conduct the study and whether the study can be conducted within the defined budget.

Clearly understanding the research problem is an essential step. Then, the team can evaluate available research design approaches, and select the research design that is most appropriate. A detailed document should be produced explaining the selected research design and the specific research methods; this document can then be used as an input for subsequent marketing research processes. The document can also contain information about the data required in the selected research design. Figure 2-12 provides sample research problems and some possible approaches to address each problem.

Research Problem	Research Design	Types of Research Design
"Why has traffic to the in-store retail channel declined over the past six months?"	Exploratory	For example, use secondary research techniques to understand whether the problem is due to a downturn in the local economy or increased competition.
"Traffic to the in-store retail channel has declined over the past six months due to increased competitor presence in the area. On what criteria can the store best compete with the new competition (e.g., pricing, product selection, or customer service)?"	Descriptive	For example, a single cross-sectional design can involve surveying current customers to understand recent changes to their shopping behavior.
"If deeper discounts are offered on accessory items, will the store see an increase in foot traffic and associated sales?"	Causal	For example, a before-after with control group study can be performed to understand the effect of advertised sales of accessory items on in-store traffic patterns.

Table 2-3: Sample Research Problems and Possible Associated Research Designs

Example of Selected Research Design:

- Sunkissed Tanning Salon is launching a spray tanning service. It is seeking to solve the business problem of negative public perception regarding traditional tanning beds. Its research problems are, “What percentage of existing customers will switch to the spray tanning service?” and “What is the target market for new customers?”

  Sunkissed will send out a survey to its existing customers to quantify the percentage likely to try the new spray tanning service.

  It will also conduct a series of cross-sectional focus groups to help understand and qualify the characteristics of the potential new target market.

3. DATA COLLECTION

Data collection is an important part of marketing research. Many significant marketing decisions are made based on the analysis of the data collected from a research project. One critical component of data collection is ensuring the quality of the data collected. Specifically, the data should be both high-quality and relevant. Data quality is the degree to which data represents the true situation. High-quality data is accurate, valid, and reliable, and it represents reality faithfully. In some instances, researchers try to obtain the same data from multiple data sources to ensure the reliability and validity of the data collected.

The following characteristics are assessed to determine the quality of data:

- Reliability—The data should be reliable such that repeating the same methods produces the same results.

- Validity—The data should measure or represent what it is supposed to measure.

Along with the quality of data, other important factors to consider in a research project are the availability of data and the affordability of the process required to collect it. Often the marketing organization already possesses enough information to make sound decisions without additional marketing research. When adequate information is not available to make a decision, additional data needs to be collected from an appropriate source. If a potential source of data exists, the researcher or the decision-maker must consider the cost of obtaining it. The data should be obtained as quickly as is required to keep the research project on schedule and at an affordable cost. If the data cannot be obtained, or if it cannot be obtained in a timely fashion, the marketing research project should not be conducted.

Researchers have the option of collecting secondary data, primary data, or both. *Secondary data* is that which has already been collected for purposes other than the problem at hand. *Primary data* is newly obtained data for a specific purpose or a specific research project.

The marketing researcher needs to decide whether to collect primary data or spend the research budget exclusively on secondary data. Researchers usually prefer to examine the utility of low-cost and readily available secondary data first to see whether they can partly or fully solve the research problem under investigation without collecting costly primary data. The source of the secondary data can be internal or external. The sources may include books or periodicals, published reports, data services, and computer data banks.

When the needed data is non-existent, outdated, incorrect or inadequate, the researcher needs to collect primary data. Most marketing research projects do include some aspects of primary data collection. Primary data may be obtained from individual consumers, subject matter experts, random samplings of a target segment, organizations, and other sources.

Depending on the source of the data required, two processes are associated with data collection. Figure 3-1 provides an overview of these processes. They are as follows:

3.1 Collect Secondary Data—This process discusses various sources of secondary data. It also discusses why and how secondary data are collected in a research project.

3.2 Collect Primary Data—This process covers various types and sources of primary data and describes how primary data are collected using various tools.

3.1 Collect Secondary Data

INPUTS
1. Senior Management Direction and Insights
2. Research Problem and Objectives*
3. Selected Research Design*

TOOLS
1. Meetings and Discussions*
2. Internal Reports*
3. Annual Reports
4. Government Publications*
5. Commercial Sources
6. General Media*
7. Bibliographic Database

OUTPUTS
1. Collected Secondary Data*

3.2 Collect Primary Data

INPUTS
1. Research Problem and Objectives*
2. Selected Research Design*
3. Senior Management Direction and Insights
4. Secondary Data*
5. Sampling Design*

TOOLS
1. Observation Techniques*
2. Experiments*
3. Qualitative Techniques*
4. Quantitative Techniques*

OUTPUTS
1. Collected Primary Data*

Figure 3-1: Data Collection Overview

3.1 Collect Secondary Data

Data collection in a research project usually starts with examining secondary data sources. No marketing research project should be conducted without a previous search of secondary sources, as there are several inherent benefits to using secondary data throughout a research project.

During the initial steps:

- Secondary data can help in understanding and defining research problems.
- Secondary sources can help define the target group.
- Secondary data can be beneficial in structuring the sample that needs to be taken.

As the research continues:

- Secondary data often provides sufficient information to answer a researcher's questions.
- Secondary data can be more accurate than primary data because of the population size under study. For example, research done by a large, multi-national company may provide more accurate and more broadly applicable data than surveys using smaller sample sizes.

When the research enters the analysis phase:

- Secondary data can help interpret primary data with more insights and validate qualitative research findings.

Two of the most significant benefits of secondary data collection is that it generally requires less time than primary data collection, and it usually generates more information for the same research budget. Sometimes researchers use secondary data because collecting primary data is prohibitively expensive or impossible.

Some of the points to be considered before collecting secondary data are as follows:

- There are many sources of secondary data available to a marketing researcher. However, the researcher should note that only a small portion of the existing secondary data may be useful.
- Secondary data may be inaccurate, outdated or unreliable. The researcher needs to evaluate the quality of data by examining its accuracy and relevance with respect to the current problem.

Figure 3-2 provides an overview of the inputs, tools, and output of the *Collect Secondary Data* process:

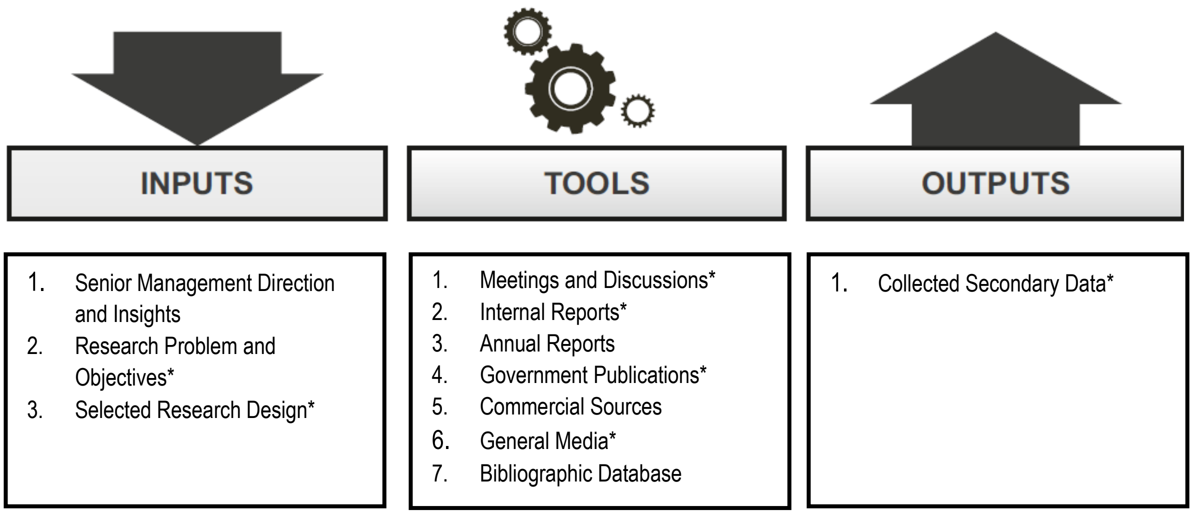


Figure 3-2: Collect Secondary Data—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

3.1.1 Inputs

3.1.1.1 Senior Management Direction and Insights

Senior management can provide useful suggestions regarding the source of secondary data and the validity of the data collected from various sources. Senior management knows if the required secondary data is available inside the organization and can authorize its collection. The senior management team should be comfortable with the validity of the secondary data and should have confidence in decisions that are made based on secondary data analysis. The decision-maker should work with senior management to determine specifics regarding the collection of secondary data such as the format of the data required.

3.1.1.2 Research Problem and Objectives*

The research problem and objectives are an output of the *Define Research Problem* process, described in section 2.1. The research problem and objectives provide a focused and definite direction for the secondary data collection process. The research problem indicates the nature of the marketing decisions to be made, which in turn dictates or specifies the kind of information that is required. The format of the required information must also be specified. Failure to provide this information will result in an unfocused and indefinite search for secondary information. While collecting secondary data, the researcher must examine and evaluate the relevance and usefulness of the secondary data to the research problem. The research problem and objectives may also indicate the possible sources of secondary data. For example, if the research problem requires investigation and analysis of nationwide data, one possible source may be a previous study conducted at the national level by a government agency. The starting point in reviewing secondary data is to make sure the research topic has been correctly identified. Identifying possible data sources involves listing all possible categories or headings under which the topic might be discussed by other researchers or authors. It is a good idea to make the topics fairly wide-ranging, so that important elements are not overlooked. Once these headings have been established, attention can then be turned to aids that will provide direction to the researcher.

Having a well-defined purpose—a clear understanding of the objectives of data collection and the kind of data that needs to be collected and analyzed—will help the researcher maintain focus and avoid becoming overwhelmed with the volume of data.

Example of Research Problem and Objectives:

- A taxi company wants to increase its sales by 100 percent in the next year. The company is evaluating various options that can help in achieving its target.

  Research Problem: Will expanding operations to Central America help increase sales?
  Research Objectives: To understand the new market and the potential to increase sales by determining the market size and competitors.
  Once the research objective is defined, the researcher will be able to finalize how to collect the data. In this case, the researcher may find secondary data such as government reports, industry reports from the transportation department, and population statistics to understand the market. The researcher may also conduct some primary research to gain a more thorough understanding of the market.

3.1.1.3 Selected Research Design*

The selected research design is an output of the *Choose Research Design* process described in section 2.2. The research design is a step-by-step plan that guides the data collection and analysis effort. In the case of secondary data collection, it might be an outline of what the final report should include, the types of data that need to be collected, and a preliminary list of data sources. It also defines the format of the information required.

Because the research design is based on the research problem, the research design is typically different for each problem. Research problems vary due to the nature of the business, the stage in the product life cycle of the product or service, the geographical location of the business, and other factors. Since the data collection methods to be used depend on the research design, it is recommended that the researcher first select the research design, then identify data collection needs, and then begin the data collection effort in order to ensure the data collection and analysis effort is focused and the results are useful.

Example of Selected Research Design:

- A university decided to open a new campus in a different city. The main objective of the research was to discover whether the university would be able to enroll a target number of students. In this case, the researcher might choose to do descriptive research, and might seek some secondary data that is readily available. Secondary data such as competitive information and population census numbers with age, gender, and education level categories might help the researcher address the key objectives of the research problem.

3.1.2 Tools

3.1.2.1 Meetings and Discussions*

While analyzing and finalizing the detailed research design with decision-makers and experts, concerned parties meet and discuss the different sources available for collecting the data required for the research project. It is important to take inputs from the decision-makers in regard to parameters such as the extent to which secondary data will be used in the current project based on budget and time allocations, their confidence in decisions made based on secondary data analysis, and the format of the data required. During these discussions, the researcher should also try to find out if any source of secondary information exists internally.

Example of Meetings and Discussions:

- In the example of the university research project described in section 3.1.1.3, *Selected Research Design*, the researcher engages the panel of decision makers and experts to understand the source of the secondary data such as the age, gender, and education of the population of the city and the number of students enrolled with other universities in the city. The panel can also provide inputs on what it feels can be additional sources of secondary data already available.

3.1.2.2 Internal Reports*

Different functions or departments within an organization have valuable sources of past data. These data or reports are generally identified in the Management Information System (MIS). Examples of internal secondary data are financial reports, sales reports, past research studies, inventory, and miscellaneous reports. Internal secondary data is not limited to the above examples and can vary depending on the industry or the research problem.

Example of Internal Reports:

- A leading food chain collects feedback from its customers regarding customer satisfaction and restaurant ratings with respect to service and meal preparation. The company wants to know the reason behind the declining sales in the last quarter. The researcher may use internal reports such as the customer service report to evaluate if the decline in sales is connected to customer satisfaction and restaurant ratings.

3.1.2.3 Annual Reports

A lot of information is readily available within the annual report of an organization. These reports provide either internal or external secondary data and can be used as a good starting point. Annual reports may have information such as sales records, marketing activity, cost information, distribution channels, and customer feedback. This data is collected by the company and is organized in accordance with the business the company conducts. Annual reports or sales data available from previous months may help the researcher estimate future projections.

Examples of Annual Reports:

- A hotel wanted to establish sales projections for the January to March quarter. The researcher analyzed the historical data in annual reports and checked the sales that took place in the same quarter of the previous years and how they correlated with the figures of the entire year's sales revenue. The historical data helped the researcher estimate future projections using statistical models.
- Annual reports can be very helpful in collecting data and targeting certain areas of a market. A review of annual reports in the investment banking industry will help to gather contacts and other data on corporate directors and other high-level professionals.

3.1.2.4 Government Publications*

Researchers may find and access huge amounts of secondary data in multiple government publications. Many governments produce reports frequently, and most of these reports can be counted on for quality and accuracy. For example, behavioral data from government sources can be more powerful and reliable than surveys and data from internal meetings or primary research. Examples of such government sources are the U.S. Census Bureau and the European Union's European Commission. Government publications or reports may include statistics on production, agriculture, import, export, population census, and expenditure surveys.

Federal, state, and local government publications or agencies can also be useful sources of information. A researcher may find a state or a local government publication that is helpful in understanding the market in depth in a particular locality. Fortunately, most governmental reports and publications can now be accessed through their websites.

Examples of Government Publications:

- A spice exporter from Cleveland seeks to estimate the export market size of its products in Brazil. The researcher studies the spice import market reports by the Brazilian government, other commercial sources, and world trade bodies such as the International Trade Centre (a subsidiary organization of UNCTAD and WTO) to gather secondary data.
- Governments will often make large amounts of industry data available to promote growth within an industry. The United Kingdom's Department of Energy and Climate Change, for example, publishes directories and interactive maps showing active areas of North Sea drilling, and identifying who operates in these areas.

3.1.2.5 Commercial Sources

Commercial sources are valuable, but usually involve a subscription to syndicate reports or membership for access. Research and trade associations remain great commercial sources for obtaining information. The information collected by trade associations is generally limited to a particular industry and is accessible to paid subscribers only. However, the information from these organizations is generally thorough and accurate. The commercial sources for secondary data will continue to be a valuable source of information.

There are many organizations that are commercial sources of external secondary data for a wide range of fields and industries. These commercial organizations allocate funding for such data collection and specialize in selling and publishing information. Market research reports and other useful publications are available from numerous organizations that charge for the information.

Example of Commercial Sources:

- Professional associations are often good sources of research data. The Canadian Society of Petroleum Geologists, for example, regularly publishes industry reports. These can be good sources of information on current industry trends, corporate direction, and future trends.
- Market research companies such as A.C. Nielsen—famous for its television ratings—provide extensive data that are ongoing and range across categories such as brand share, promotion, pricing, sales volume, audience size, and consumer behavior.

3.1.2.6 General Media*

General media such as the Internet, online databases, journals, broadcasts, and print media are major sources of external secondary data. The Internet proves to be the easiest and the quickest source of secondary data. A researcher can access a variety of websites, journals, and online versions of traditional print media to obtain required data. Archives of newspapers can be accessed online, making them useful resources for the researcher. Online databases are valuable sources of information because they usually contain the most up-to-date data, provide fast search capabilities, and are low-cost and convenient.

Examples of General Media:

- The Financial Times, The Economist and The Wall Street Journal are trusted valuable sources of information on market trends and the economy.
- General media serves to be a major source of information for organizations in their marketing decisions about topics such as advertisements and promotions. For example, sources such as the United States' Federal Communications Commission, the UK's Ofcom, newspapers, and magazines provide information regarding the most watched television programs. Organizations use this information to broadcast their advertisements during particular programs that are likely to help each organization achieve maximum benefit from its marketing efforts within the target segment it shares with a program.
- In sports, major tournaments can attract a huge number of viewers across the globe. Previous research, trends, and broadcast information help researchers by providing key information, such as the number of viewers, their demographics, and their psychographic characteristics. A leading sportswear brand might use this information to make decisions regarding tournament and apparel sponsorship and to determine the value of broadcast advertisements during live matches as a brand-building activity. These channels can provide a captive audience for the sports apparel brand given the audience's interest in sport.

3.1.2.7 Bibliographic Database

Bibliographic databases are generally composed of citations of journal articles, print media such as newspapers and magazines, and marketing research reports. The data obtained from these sources often provide summaries or abstracts. Journals and periodicals are some of the most important sources for reliable data. They provide information that is up-to-date and on a specific topic or issue, which books cannot provide.

3.1.3 Outputs

3.1.3.1 Collected Secondary Data*

The collected secondary data contains information collected from both internal and external sources. The overall direction, accuracy, and success of the marketing research project depend on the data collected. Data collected inaccurately or incompletely may lead to incorrect research results, poorly planned marketing activities, and an organization's eventual failure to achieve its objectives. In some situations, the collected secondary data may be outdated or may be less relevant than primary data. The collected secondary data can be used as an input for other processes in Marketing Research including *Data Processing and Analysis*.

Examples of Collected Secondary Data:

- GeoPro is conducting a research on offshore drilling. The company reviewed all available governmental data to determine active players in the target area. It then captured management information from the appropriate annual reports and associated corporate websites.
- A telephony hardware company selling primarily to call center companies wants to expand its sales focus to the banking and financial service industry. To better understand the telephony and integration needs of this target market, the company purchases research reports from a well-known analyst firm.

3.2 Collect Primary Data

When a researcher collects new data for a specific research project, the data is considered primary data.

The research requirements will vary from project to project, with many research projects requiring both primary and secondary data to solve the research problem. Some research projects can be solved with the sole use of existing secondary data, while in other cases no secondary data exists and the research project can only be solved with the use of primary data. The research needs, along with existing resources and scheduling requirements, dictate the costs of the project.

As a rule, a researcher should always try to collect and analyze secondary data before moving to the collection and analysis of comparatively costly and time-consuming primary data. In some cases secondary data may be inadequate or unusable. When the needed data do not exist or are outdated, inaccurate, incomplete, or unreliable, the researcher needs to collect primary data.

Primary Data Sources

There are various sources of primary data. Primary data may be obtained from individual consumers, subject matter experts, random samplings of a target segment, and various organizations, among others. Some of the major sources of primary data are as follows:

- **Internal Organization**—The organization is not only a good source of internal secondary information, but also it can be a good source of primary data. Interviewing and surveying relevant employees in the organization can be a good starting point for primary data collection. For example, the internal sales force is a good source of primary data in a research project that involves forecasting future sales.

Example of Primary Data Collection from the Internal Organization:

- Most companies are interested in keeping employees happy because studies show that engaged employees are more productive and are better contributors to their employer's goals and objectives. Employee satisfaction surveys are done to evaluate the satisfaction of employees on key parameters such as satisfaction with senior management, immediate managers, and company policies.

  Employees are also approached if improvements are required for a particular process, or if a new process needs to be developed to address any existing issues.

- **External Environment**—The external environment includes all types of users, customers, potential customers, and any external entities, such as competitors, that can influence or be influenced by the outcome of the research project. Users or customers are important first-hand sources of information related to customer demands, views, beliefs, and intentions.

Example of Primary Data Collection from the External Environment:

- Every company wants to understand the impact of its marketing activities on its target audience and, therefore, has to conduct surveys among its customers. A leading apparel company tests its creative concepts with its target audience before launching them into the market so that it is confident in the message the company is conveying through advertisements.

  After the launch of the campaign, the campaign is evaluated again with feedback from the target audience to determine its impact on the brand image.

- **Distribution Channel**—Another important source of primary data is the distribution channel. Vital information can be collected from wholesalers, retailers, manufacturers, and suppliers. In fact, the role of wholesalers and retailers has increased tremendously in recent years in many product categories where customer decision-making is greatly influenced by the channel partners.

Example of Primary Data Collection from the Distribution Channel:

- Retailers and distributors provide the distribution channel for a company to reach its customers. Consumer packaged goods companies, where the differentiation among the products is low, conduct a lot of research among distributors and retailers to understand the push they provide for their brand in the market. The retailers act as influencers when consumers are deciding which brand to buy, and their push can promote the brand in the market. Therefore, it is necessary for the company to understand their needs and perception about the brand.

Types of Primary Data

There are various forms of primary data. Some common types are as follows:

- **Demographic Data**—Demographic data are related to characteristics such as the gender, age, income, education, occupation, marital status, ethnicity, and social status of the target group. Demographic data are important because it helps marketers profile a target group. Demographic primary data help group the respondents into consumer segments. For example, a soft drink brand may find that the preferences of people in the 25–35 age group are different from the preferences of people in the 35–45 age group. Data can also be categorized by life stage including early childhood, youth, young adult, newly married, married with young children, married with teens, empty nesters, elderly, and retired.

- **Psychographics and Lifestyle Data**—This kind of data is related to personality traits, interests, lifestyle, values, and opinions of the target respondents. Marketers often combine psychographics and lifestyle information with demographic information to obtain an important perspective of the target market.

- **Intentions**—Intentions refer to the anticipated future behaviors of an individual. This is a subject of interest to marketers who want to solve a research problem related to future consumption rate or demand. Although there will always be some differences between consumers' intentions and actual practices, sellers believe that expected future behavior is a useful indicator to evaluate several possible alternative offerings.

- **Attitudes**—Attitudes refer to a person's feelings, convictions, or beliefs toward an object, idea, or an individual. Since attitude impacts behavior, it is of great importance to marketers. It is difficult to measure or observe behavior in all conditions, and measuring attitudes of the respondents can provide a good indication of possible behaviors.

- **Awareness/Knowledge**—This data refers to what subjects do or do not know about an object of investigation. Information influences behavior and marketers often want to know how the behavior of customers changes with their level of awareness regarding a particular product, brand, object, or industry. This information helps marketers determine the image, experience, and feelings of consumers who are familiar with a product and make distinctions between consumers who are familiar with a particular product or industry and those who are not.

- **Motivations**—A person's actions are the reflection of his or her inner state. Marketers often want to know the motives that direct specific consumer behavior. When a respondent is asked about the factors that are likely to influence his or her decision, a researcher is trying to identify the motives that influence behavior. Motivations can include users' category, brand-purchasing motives, value systems, and perceptions among others.

- **Behaviors**—Behaviors are the actions taken by respondents. Data related to consumer behavior is of great importance to marketers. Questions regarding respondents' behaviors toward a particular situation can be asked to them directly and can be included in a survey. However, the responses may not represent the actual behavior of the respondents. Observation techniques are more often used to understand the actual behavior of respondents. To further explore respondent behavior, marketers can categorize consumers according to product usage and user status. For example, transactional or research data can be used to divide a user base into cohorts by units consumed or dollars spent (e.g., light, medium, or heavy). Purchase behavior is also an important factor; when considering purchase behavior, marketers might categorize consumers as non-users, potential users, first-time users, regular users, or former users.

Primary Data Collection Methods

Primary data can be collected by a process of observation, qualitative research, or quantitative research, and these approaches can be used separately or in combination.

- **Observation Research** involves the collection of information with regard to the behavior of individuals, objects, and organizations without any questions being asked of the participants.

- **Qualitative Research** uses both structured and unstructured approaches to obtain unquantifiable insights into behavior, motivations, and attitude from a small number of selected individuals.

- **Quantitative Research** uses structured approaches to obtain measureable insights into the behavior, motivations, and attitudes of target respondents by involving a sample of the target population.

Depending on the type of research being conducted, this process is also likely to involve the development of data collection forms, the determination of the sample of respondents to take part in the research, and the actual collection of data.

Figure 3.3 provides an overview of the inputs, tools, and outputs of the *Collect Primary Data* process.

INPUTS	TOOLS	OUTPUTS
1. Research Problem and Objectives* 2. Selected Research Design* 3. Senior Management Direction and Insights 4. Secondary Data* 5. Sampling Design*	1. Observation Techniques* 2. Experiments* 3. Qualitative Techniques* 4. Quantitative Techniques*	1. Collected Primary Data *

Figure 3-3: Collect Primary Data—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

3.2.1 Inputs

3.2.1.1 Research Problem and Objectives*

The decision to collect primary data for a research project is a very important one. The research problem and objectives not only influence the decision to use primary research but also provide insights to the researcher regarding the application of collected data and the type of primary data that needs to be collected. Once the decision is made, it is important not to lose sight of how a researcher arrived at the point of acknowledging the need for primary data. The primary data collected should be in the form needed to answer the question related to the problem, test a hypothesis, or in some way contribute to improved decision-making. For example, from the research problem and objectives, the researchers will know that they need to measure consumer preferences if consumer preference is one of the key criteria identified by the decision-makers. A particular type of primary data collection should be included in the research project only if there is a reason for its inclusion, ultimately linking the type of data to the solution of the research problem.

While designing the primary data collection process it is important to decide the type of methods to be used, the data to be collected, and the ideal sample size, all of which are derived from a good understanding of the research problem and objectives.

Example of Research Problem and Objectives:

- A university wants to improve its ranking at the international level and wants to double student enrollment in several select programs in the following term. The university wants to improve its current programs in order to attract good-quality students from around the world.

  According to the requirement of the university, the researcher defines the research problem as "Evaluate the current teaching methods and the performance of the faculty." The objective of the research is to identify competent faculty and to compare the current methods of teaching against the most effective pedagogical practices. This will enable the university to reward the competent faculty and to create high-quality programs.

  In this project, the researcher can gather secondary data about course enrollment and faculty details from the university's internal records. Data on the best pedagogical models can be collected from various books, journals, past surveys, articles, and other online or offline resources. However, to evaluate the current teaching methods of the faculty and their popularity as defined in the research problem and objective, the researcher needs to collect primary data. The researcher may opt to create a survey questionnaire that asks students to rate faculty on various parameters and provide their views on the teaching methods (although results collected from non-experts will have very little value, and, if useful at all, will require expert interpretation). The researcher can also conduct an expert panel discussion to identify best practices in teaching.

3.2.1.2 Selected Research Design*

The selected research design is the master plan for conducting the research project. It provides the overall guidelines on how the researcher can obtain answers to the research questions and test hypotheses. The research design also includes the specific methodology that should be used to collect the data. The research design includes information regarding the use of communication and/or observation approaches to obtain the data, the degree of structure and approach of the research, and the process of administering the research. The research design will also facilitate the consideration of ethical and legal issues while designing primary data collection processes so that no legal issue arises while collecting the data.

3.2.1.3 Senior Management Direction and Insights

While analyzing and finalizing the detailed research design with the senior management, decision-makers and experts, concerned parties meet and discuss the different sources of data collection required for the research project. If the researcher feels the need to collect primary data, he or she should inform the decision-maker and discuss the techniques that can be used to collect this data. It is important to listen to feedback from the decision-makers while finalizing the scope of collection of primary data under given budget and time constraints. The senior management and decision-makers should also be consulted on the type of primary data to be collected, the intention behind collecting particular types of data, the suggested method and tool to collect the required data, the preferred format of collected data that will help in addressing the problem at hand, and so on. During the discussions, the researcher should also try to find out possible internal secondary data sources.

Experts can provide valuable input regarding the preferred method of data collection, types of data, format of data collection forms, sample size, and more.

Example of Senior Management Direction and Insights:

- A real estate company wanted to understand the requirements of its target buyers for a villa township located in the outskirts of the city along with the price consumers would be willing to pay for each additional amenity that they list as their requirements. The researcher designed a research project where in-depth interviews would be conducted with high net-worth individuals who would be able to afford such a place and could give honest feedback on the price that they would be willing to pay for each amenity.

  During the meeting to discuss the research design, senior management suggested interviewing couples instead of only individuals because usually decisions related to real estate are made jointly by both husband and wife. As a result, the research design was modified to incorporate couple interviews rather than individual in-depth interviews.

3.2.1.4 Secondary Data*

The researcher should always start by examining secondary sources of information. In order to avoid wasting resources and time, the researcher should avoid collecting primary data if someone else has already collected the required data.

Before collecting primary data in a project, the researcher needs to determine the amount and type of information that still needs to be collected through primary research. If the researcher already has access to high-quality secondary data that partially addresses the problem at hand, he or she may consider collecting only the additionally required information through primary data collection in order to solve the problem. Comprehensive analysis of the secondary data forms a vital foundation and helps the researcher to decide on a more focused approach for primary data collection. The decision-maker and researcher can use the ideas generated from secondary data as a strong foundation for designing the primary data collection effort.

Secondary data can also be helpful in the interpretation of primary data by enabling researchers to see the broader picture of what the data means in the context of other current or historical developments in the market. It can provide a source of comparative data for the data gathered from primary research.

Example of Secondary Data:

- A financial institution wants to launch a new product in the market. The company does not have any information on the current status of various financial products that are available in the market, provided by the competition. It wants to launch a product that is differentiated and has an immediate demand among buyers.

  Thus, the company first decides to do secondary research to discover the various financial products present in the market so that it can plan the options of its future product accordingly. This data would be used to plan the product concepts to be tested in the market using descriptive research techniques.

3.2.1.5 Sampling Design*

While sampling design is a part of research design, the topic warrants further discussion in the context of defining primary research requirements, as sampling design has specific application requirements that should be considered before choosing an approach.

A population is the entire group of objects having characteristics of interest under study. The subset of a population that is chosen for the study is known as a “sample.” Choosing the portion or subset of a population is known as “sampling.” The chosen sample must represent all or most of the features of the population from which it is chosen. To ensure that the chosen sample appropriately represents the population, a strategy is required. This strategy is known as a “sampling strategy.” The sampling strategy is a plan or strategy created to make sure that the sample of the population on which data will be collected is accurately representative of the group identified for study.

The task of sampling is undertaken when information regarding a process or product is not readily available, and analysis of the entire population on which the critical information is required is not feasible or possible (i.e., such an undertaking would be too time-consuming and too costly). Because sampling reduces costs and employs fewer human resources (among other benefits), it is commonly employed in most industries that require critical information regarding a process or product.

Sampling is also used when the data collection is a destructive process. For example, CDF Inc. is a mineral water manufacturer that produces bottled mineral water. The quality assurance team tests the quality standard of the mineral water by randomly selecting a sample of bottles taken from each production batch. In the testing process, they open the bottles and introduce chemicals into the contents, thus destroying the sample. These bottles will no longer be hygienic enough for sale and the water will be contaminated; testing the entire population of bottles would result in no revenue for the company, and therefore a sample is tested.

Types of Sampling Strategies

Researchers can choose from a number of different types of sampling strategies. The type of strategy chosen should appropriately suit the research objectives. The selected sampling strategy impacts the amount of data that can be collected and the margin of error that exists when it comes to the generalization of research results. The margin of error is typically a small amount that is allowed for in the event of miscalculation.

Sampling strategies are classified as either probability sampling or non-probability sampling.

1. **Probability Sampling Strategies**—Probability sampling strategies are the most reliable sampling strategies because the margin of error is minimal due to the statistical procedures used. In these strategies, every component in the population has an equal and independent opportunity to be chosen.

   The four main methods of probability sampling are simple random sampling, systematic sampling, stratified sampling, and cluster sampling.

   - **Simple Random Sampling**: This is the most basic type of probability sampling in which each individual is chosen randomly, so that each individual has an equal and unsystematic chance of being selected at any stage.

     Simple random sampling can be conducted in the following ways:

     - **Random numbers**: The best way to conduct random sampling is by using random numbers generated by computer or using a table of random numbers. For example, from a population of

1000 employees in a company, 50 employees are to be selected randomly for conducting a marketing campaign.

> *Example of Random Sampling:*
>
> - Random Sampling can be done using a function =*RANDBETWEEN(bottom, top)* in an Excel worksheet. Assign each employee a number between 000 and 999 (both inclusive). In the worksheet, write the formula as =*RANDBETWEEN(0, 999)*. Copy this formula and paste it in another 50 cells. This will generate a random sample of size 50.

- **Lottery method**: This is a very popular method of selecting a random sample from a small population. In this method, each element of the population is assigned a name or number. Each name/number is written on a paper slip and folded to look identical to the other slips. All the slips are then deposited into a box and mixed thoroughly. The required number of slips is drawn randomly from the box. One drawback of this method is that it is not feasible when the population size is large. Another drawback could be that if the slips are not mixed properly there may be some pattern in the chosen sample.

- **Systematic Sampling:** In systematic sampling samples are selected according to some specified systematic rule, such as the selection of elements from the population from a random starting point and at uniform intervals.

  For example, a sample of 10 questions is to be selected out of a set of 100 questions by using the systematic sampling technique. Assuming the questions are numbered 1 to 100, a starting point is selected randomly from the first 10 questions (i.e., questions numbered 1 to 10). Thereafter, every 10th question is chosen. If the random starting point is question number 7, then the selected question numbers are 7, 17, 27, 37, 47, 57, 67, 77, 87, and 97.

  The interval length is the nearest integer value of the ratio of population size to sample size.

  In this example ,

  $$\text{Interval Length} = \frac{Population\ Size}{Sample\ Size} = \frac{100}{10} = 10$$

  The major drawback of systematic sampling is that the systematic rule may match some hidden trait of the population and introduce bias in the sample. Systematic sampling is applicable only if the population under study is homogeneous. Also, it must have a degree of randomness so that the sampling interval chosen does not hide any pattern.

A significant difference exists between simple random sampling and systematic sampling. Each unit of the population has an equal probability of selection in both cases but each sample does not have an equal probability of being chosen in systematic sampling unlike simple random sampling.

- **Stratified Sampling:** In this sampling technique, the entire population is divided into relatively homogeneous groups. These groups are known as strata. A random sample is chosen from each stratum. The size of the sample corresponds to the proportion of that stratum in the population as a whole. Alternatively, equal numbers of elements are drawn from each stratum and a weight is given to the results according to the stratum's proportion of the entire population.

- **Cluster Sampling**: In this sampling technique, the total population is divided into groups or clusters. These individual clusters are essentially similar to each other. A random sample of these clusters is selected.

Both stratified and cluster sampling requires the entire population to be divided into some well-defined groups. However, there are some significant differences between these two types of sampling. With stratified sampling, statistical analysis is conducted on elements within strata, and a random sample is chosen from each of the strata. Conversely, with cluster sampling, since each cluster is treated as a sampling unit, analysis is completed on a population of clusters, and only randomly selected clusters are studied.

Stratified sampling is used when each group has a small variation within itself but there is a wide variation between the groups. Cluster sampling is used when there is considerable variation within each group, but the groups are similar in nature.

2. **Non-probability Sampling Strategies**—Non-probability sampling strategies are not as reliable as probability sampling strategies. The selection procedures in these strategies involve non-random methods. As a result, the subjects in the population do not have an equal chance of being selected as part of a sample. These types of sampling strategies are less likely to produce representative samples than probability sampling strategies. Regardless of this factor, many researchers have successfully used and continue to use these strategies. The three main strategies of non-probability sampling are Convenience, Quota, and Purposive.

   - **Convenience sampling**: Convenience sampling is widely used in student research projects. In convenience sampling, elements that are easy to get are chosen from the population for the study. Thus, this strategy may lead to biased data.

Examples of Convenience Sampling:

- Industry trade shows can offer the opportunity for convenience sampling of attitudes about a company's new product or service. The characteristics of participants in the sample may be beyond the company's control, so the results may or may not be representative of the population as a whole.
- One classic example of convenience sampling is an opinion poll for the 1936 U.S. presidential election conducted by a popular weekly magazine of the time. The magazine erroneously predicted a victory for Republican candidate Alf Landon with the incumbent Franklin D. Roosevelt receiving 43 percent of the votes. The magazine created a list of ten million people by combining phone books and lists of automobile club memberships. The magazine then sent a questionnaire to the ten million people, and 2.4 million people returned the questionnaire. Of the respondents, 43 percent planned to vote for Roosevelt and the magazine predicted the poll result based on that figure. In reality, Roosevelt decisively won the election with a whopping 62 percent of the votes, carrying every state except for Maine and Vermont. The magazine's forecast was incorrect because it was based on a highly non-representative sample of the electorate—mostly car and telephone owners, as well as the magazine's own subscribers—which underrepresented Roosevelt's core constituencies. Use of convenience sampling here led to biased data and an incorrect prediction.

- **Quota sampling**: Quota sampling is a non-probability sampling technique where the assembled sample has the same proportion of individuals as the entire population with respect to known characteristics, traits, or focused phenomenon. In quota sampling, a population is first segmented into mutually exclusive subgroups and then judgment is used to select the elements from a subgroup based on a specified proportion. For example, an interviewer surveying the first fifty women wearing black dresses that he or she meets on a particular day is using quota sampling.

  Quota sampling should be used only as a last resort because it is prone to the selection of a sample that does not precisely reflect the characteristics of the population, and thus introduces bias into the study.

Example of Quota Sampling:

- In order to determine competitor pricing of a similar product, an online search may be conducted. The first five similar products may be used as a benchmark. This would be a quota sample and may not be accurate, though it will still provide a base level of data from which to proceed.

- **Purposive sampling**: A purposive sample is one that is selected based on the knowledge of a population and the purpose of the study. The main goal of purposive sampling is to focus on particular characteristics of a population. The subjects are selected because of the characteristic under consideration. Purposive sampling is applied only in situations where there are no other alternatives due to difficulties in locating and recruiting the desired population sample for the study.

Examples of Purposive Sampling:

- Sales representatives may conduct purposive sampling with clients who have discontinued service. This group is targeted due to this trait that they share—having discontinued their service. Attitudes regarding industry trends, technological advancements, competition, and pricing can be gathered.
- A consumer durables company that is conducting a competitor analysis may want to interview only those people who use a competitor's product.

3.2.2 Tools

3.2.2.1 Observation Techniques*

Observation techniques are considered valuable because they can provide insights on the consumer's actual behavior. Observation techniques are used by researchers to record consumer behavior either with or without their knowledge. Types of observation techniques include structured observation, unstructured observation, obtrusive observation, unobtrusive observation, and ethnographic research.

1. **Structured Observation**—Structured observation, also referred to as systematic observation, is a technique for data collection in which the researcher collects the information directly without the mediation of respondents, interviewees, and so on. In this structured technique, the data is collected according to a set of predefined rules.

Example of Structured Observation:

- A retail chain can use paid observers to conduct reviews of the shopping experience at a given store. These observers, often referred to as mystery shoppers, will behave as though they are regular customers. Their goal will be to evaluate the shopping experience based on predefined criteria. This gives the retail chain data on the quality levels of the staff, presentation quality of the store, and overall shopping experience. This data can then be used to improve the customer experience and help to improve brand image and reputation.

2. **Unstructured Observation**—In unstructured observation, the researcher has some general ideas of what might be relevant, but not of what specifically needs to be observed. Therefore, observation is unstructured, holistic, and unfocused, as the researcher attempts to record as much as possible about the participants and the setting in order to determine themes of interest.

Examples of Unstructured Observation:

- The sales team for a home furnishing brand is attending a trade show. In addition to presenting their own products to attendees, the staff will visit the booths and demonstrations of their competitors. This gives them a chance to observe the marketing strategies and presentation methods that the competition is employing and discover customer reactions to the competition's marketing efforts. The criteria and desired outcomes are not defined prior to the review. All the observations and insights gained will be documented and compared at the end of the trade show.
- A researcher for an airline company is trying to understand the difficulties that passengers face while traveling with the airline. The investigator will travel on the airline and observe the passengers to identify any points of discomfort that the travelers experience during the flight.

3. **Obtrusive Observation**—These are commonly used within a sample group to understand consumer reaction to a product, commonly referred to as obtrusive technique. In this scenario, the consumer or the sample is aware of the observation.

Examples of Obtrusive Observation:

- A statistical research company conducting research on behalf of a radio station may send out logbooks to selected households. The members of the household will be asked to complete daily records of their radio listening habits. This is time consuming for the population being observed, but the results give an extremely accurate picture of listener behavior.
- A car manufacturer may bring test drivers to drive a model in various conditions. Researchers will examine the performance of each test driver and record their comments.

4. **Unobtrusive Observation—**When the consumer is not aware of the observation, the technique is referred to as unobtrusive observation.

Example of Unobtrusive Observation:

- A call center for a cable company might test market a new corporate slogan by introducing the slogan at the start of each phone call. Customer reactions to this slogan can be recorded and reviewed to test the marketing. The customers may be aware that their phone calls are being monitored for quality purposes, but they are not aware that their reactions to the new slogan are being measured.

5. **Ethnographic Research—**This type of research is conducted by closely examining a customer's behavior towards a certain product over a period of time. This type of observation is exploratory and can

be valuable for understanding the perceptions toward a product or service over time. It is normally considered honest.

Example of Ethnographic Research:

- A leading beer manufacturer interested in understanding the drinking behavior of customers observed several pubs, restaurants, and parties over a period of time to understand changing trends and attitudes toward particular types of beer. The manufacturer even accompanied the owners while ordering or purchasing new stocks.

All observation techniques can help researchers discover important insights about a wide variety of consumer behaviors and attitudes toward certain objects, brands, trends, and industries.

3.2.2.2 Experiments*

Experimentation is often the method used by researchers to gather primary data. The causal research method is normally conducted using experiments. Here, the researcher alters one or more variables while observing the changes or effects on another variable. For example, in a retail store a researcher may choose to alter price, design, shelf space, positioning, and advertisement while observing the effect on the sales of the product.

While conducting an experiment, the researcher identifies the key variable that can possibly affect other variables. This method of collecting primary data can be impacted by other variables unless most of the factors are considered as constants, which is highly impractical. For example, an increase in the advertising budget can increase sales, but there are certain external factors, which can also affect sales, such as economics, weather, and competitors' actions. The external factors always play a major role and are addressed (i.e., controlled for) by conscientious researchers in order to increase data reliability and validity.

Experimental design and the validity of experiments are detailed in section 2.2.2.5.

Example of Experiments:

- A leading soft drink manufacturer lost the majority of its market share to competitors. Management wondered if an increase in the volume of the drink per unit for the same price would increase sales. The soft drink company designed an experiment where the marketing mix stayed the same, but the size of the drink changed from small to large. Due to the change in packaging and SKU (Stock Keeping Unit) size, the company was able to increase sales. Providing the soft drink in a larger volume resulted in increased costs for the manufacturer; however, the additional costs were more than compensated for by the additional revenue.

3.2.2.3 Qualitative Techniques*

Qualitative techniques are also sometimes referred to as the “soft side” of marketing research. A variety of factors related to the research problem cannot be studied through quantitative research and require an understanding of softer aspects such as thoughts, feelings, emotions, behaviors, and so on. Many researchers practice a pluralistic approach, which involves combining qualitative and quantitative research techniques, thus taking advantage of the benefits of both types. Among the various qualitative techniques used by researchers are in-depth interviews, projective techniques, and focus group discussions.

1. **In-depth Interviews**—These interviews are normally open-ended, one-on-one interactions conducted either telephonically or face-to-face, and designed to collect individualistic behavior dynamics. Although these interviews are unstructured, the questions and probing are structured around the research problem.

   The researcher should be able to gather the data, analyze it, and write a report at the end of the interview. In-depth interviews are not suitable for all types of research problems; they are typically suited for the following research scenarios:

   - when there is a deeper need to understand individualistic behavior
   - when the research involves sharing personal or confidential information or discussing embarrassing topics
   - when the research involves interviewing professionals or experts (as it is often more convenient and feasible to interact with them in a one-to-one manner)

Example of In-depth Interviews:

- A consumer packaged goods company wants to understand more about the purchase behavior of its target consumers. The company conducts in-depth interviews with a sample of fifty people. The following questions are asked:
  - How do people become aware of their need for the product?
  - How do consumers find the product?
  - How do consumers make their final selection?
  - How do consumers order, purchase, and pay for the product?
  - What happens after they buy the product?
  - How is the product used and prepared?
  - What are consumers really using the product for?
  - What do consumers need help with when using the product?
  - How do consumers get help?
  - Do customers review the products? How? What do they say? Where do they say it?
  - How is the product refunded or replaced?
  - What happens after consumers finish using the product?

2. **Projective Techniques**—The projective technique is a questioning technique used to obtain indirect responses underlying motivations, beliefs, attitudes, or feelings regarding issues of concern. The main objective of the projective technique is to discover and interpret consumers' perceptions of a product or service that they might not express directly.

   There are different types of projective techniques:

   - **Association**—Subjects are provided with a hint or stimulus and asked to respond with the first thing that comes to mind. Word association is a common example of the association technique.

   - **Completion**—Subjects are provided with an incomplete stimulus situation and asked to complete it in any way they wish with the constraint that the completions meet certain standards of rationality or form. For example, a story completion may require a chronologic order.

   - **Construction**—Respondents are required to go beyond simple association and to consrtuct or create a more elaborate product, which is usually a complete art form, such as a story, description, dialogue, or picture. In a construction technique, the initial structure provided to the respondents is less than is the case with a completion technique. For example, the researcher might ask respondents to create a collage of what they feel when they experience a product, service, or concept under study.

   - **Choice or Ordering Techniques**—Choice or ordering techniques require the respondent to select from a set of alternative arrangements that fit some specific criteria, such as relevance, attractiveness or meaningfulness (e.g., a picture arrangement test).

   - **Expressive**—In expressive techniques, respondents are given a visual or verbal situation and asked to relate the attitudes and feelings of other individuals to the situation (e.g., a role play).

The projective technique has some limitations. Using this approach can be time-consuming and expensive. It requires a smaller sample size, has a high non-responsive rate, and can be misinterpreted. Interpretation of the data collected from a study that uses the projective technique can also be time-consuming and complex.

Example of Projective Techniques

- A deodorant brand plans to determine its brand image, which can be associated with the perceptions of the consumer. Brand image is the total brand personality perceived by the customer. Using the construction projective technique, the brand asked respondents to create a collage of what they feel when they experience the deodorant under study. Based on feedback, the brand was able to characterize its image as a young, successful man who is handsome, pleasant, and enjoys the outdoors.

3. **Focus Group Discussions**—A focus group is used to collect qualitative data from five to ten respondents by conducting a small meeting in the presence of a moderator. Instead of collecting responses to specific questions, this meeting aims to create a forum for informal discussion about specific products or certain topics. The main objective of the focus group is to acquire insights into customer preferences and behavior. The informal and relaxed atmosphere often helps uncover unexpected insights about the product. Focus groups are usually conducted for exploratory research where there is a need to understand a complete picture of market dynamics.

   Focus group participants are carefully selected on the basis of common traits such as similar buying behaviors, demographics, and socio-economic factors. Homogeneity among the participants prevents conflicts on issues that are not relevant to the discussion. One of the key criteria for selection can also be familiarity of the participants with the products or topics being discussed. For example, the participants of a focus group formed to evaluate the use of a new mobile application for smartphones should have adequate experience in using smartphones. Focus group size is usually kept small enough for the participants to have enough time to share their opinions and yet large enough to gather diverse views about the topic. Having too small of a group can limit the information that can be gathered, and too large of a group can result in a lack of opportunity for all participants to share their insights. Questions or topics for the focus group need to be developed based on the marketing research problem and objectives. When based on the research problem and objectives, the role of each focus group will also be clearly defined. The type of questions used to initiate discussion should be open-ended. For example, questions such as "What do you like about the product?" and "What would your suggestions for this service be?'' would be appropriate for a focus group discussion. Questions that invite a "yes-or-no" response are generally not used in focus groups.

   The moderator of a focus group plays a vital role in initiating and leading the discussion. The moderator should have characteristics similar to the participants and should have experience in handling groups. She will guide the discussion through the topics and ensure that the discussion aligns with the objectives. Even experienced moderators can sometimes find it difficult to manage participants who have a dominating personality and tend to disrupt the discussion. A skilled moderator will be flexible enough to alter the flow of discussion amid such disruptions. During lengthy discussions, the moderator should ensure that adequate arrangements are made to provide participants with a comfortable environment to relax and provide insights to discussion topics.

   In most focus groups, audio and video of the discussions are recorded for analysis and reporting purposes.[9] Notes and comments can be gathered by the moderator, which will help in creating the summary of the discussion. The report of the focus group discussion should be descriptive in nature. Instead of just summarizing the data, reporting should also provide a detailed interpretation of the data.

[9] Naresh.K.Malhotra, David.K.Birks, *Marketing Research – An Applied Approach* (2006).

Examples of Focus Group Discussions:

- Focus group research is used extensively by pharmaceutical companies to test concepts related to new drugs and marketing of over-the-counter (OTC) and prescription products. Using focus group discussion, pharmaceutical companies explore consumers' and physicians' opinions about various marketing mix elements such as product concepts and packaging, advertising, and promotional ideas. Focus group research will help the pharmaceutical company gain physicians' insights on various important matters and opinions, such as the types of users that will find a drug most appropriate, how important the drug is perceived to be in the treatment of patients, and when and how the physicians will prescribe the drug to patients. Focus group discussions will also help to gather input from physicians and end-users on the preferred positioning for the product or service. Focus groups can help in identifying the information needed to target customers for the new product. If advertising and merchandising materials have already been created, a focus group will provide information if the target customers (physician and/or retail consumer in this case) relate well to the advertising or merchandising materials aimed at them. Because many of drugs are not supported by mass media advertising, focus group discussions will also enable the company to gather useful information about packaging, which is helpful especially with OTC drugs. Focus groups also help pharmaceutical companies understand the opinions of physicians and end users regarding the service they expect to receive from the company thereby helping the company in creating customer service and loyalty programs.
- A dental clinic conducted a focus group discussion to improve the process for patients arriving at the out-patient office. The focus group was composed of patients who had experienced the clinic's facilities, and the collated feedback was provided to the clinic management. Based on the recommendations of the focus group, the waiting area layout was improved and clear, easy-to-follow instructions were provided to first-time patients to avoid any inconveniences. A second focus group was conducted a few months later to evaluate the success of the implemented changes.

3.2.2.4 Quantitative Techniques*

Quantitative data collection techniques depend heavily on the sampling methodology and the data collection instrument used. These techniques are concerned with testing hypotheses derived from theory or estimating the size of a phenomenon of interest.

One research or data collection instrument used in these techniques is the questionnaire, which is a set of pre-determined questions that can elicit both subjective and objective responses. There are a number of factors to consider in designing a questionnaire, including the characteristics that will be measured in the questionnaire, the scale on which the characteristics will be rated, the types of questions to ask, and the set of responses that will elicit meaningful data.

Designing an appropriate questionnaire that will elicit high-quality, analyzable data requires careful consideration of all of these factors. One of the first steps in designing a questionnaire is to determine the characteristic that needs to be measured and then determining an appropriate scale for rating the characteristic.

Measurement is the act of assessing the level of presence of certain characteristics in the variable under consideration. The characteristics measured are attributes or qualities of the variable. Characteristics that a researcher might measure include certain demographic characteristics, consumer perception of quality or user friendliness, or the reasons for a decrease in sales. Each characteristic is measured using a scale that will enable respondents to rate the attribute of the variable. For example, a restaurant may ask patrons to rate their dining experience (the attribute measured) on a scale of poor, fair, good, or excellent.

Types of Scales

Scales can be classified into "Categorical" and "Metric" scales.

- **Categorical Scales** measure variables with a few distinct levels of measurement for values. For example the taste of a newly launched beverage can be measured as sweet, bitter, or bland. Two types of categorical scales are Nominal and Ordinal scales.

- **Metric Scales** measure variables using numbers on a continuous scale. Rating a new book on a scale of 1 to 10 is an example of a metric scale. Two types of metric scales are Interval and Ratio scales.

The level of measurement or type of scale chosen is a critical concept to consider in the design of a questionnaire for two reasons. First, the chosen scale will determine the amount of information to be derived from the scale. The amount of information derived is lowest when the nominal scale is used and highest when the ratio scale is used. Second, the chosen scale also determines the statistical analysis possible for a particular scale. Figure 3-4 shows the types of scales available to researchers.

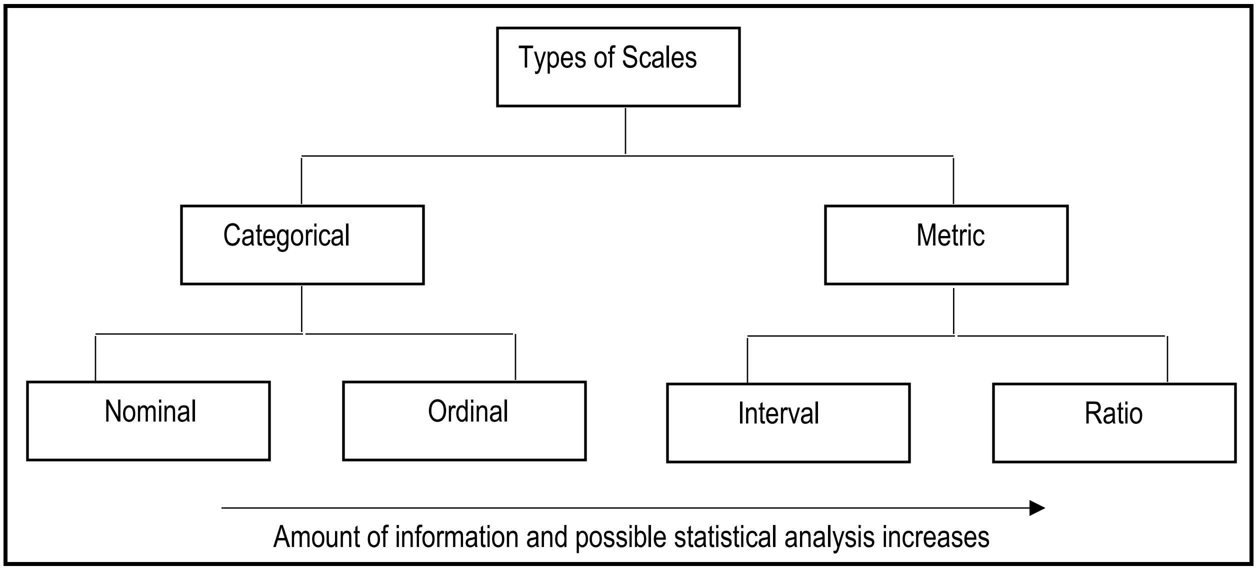


Figure 3-2: Types of Scales

1. **Nominal Scale**—A nominal scale is used for labeling variables. It is the simplest form of scale and can only be used to carry out the simplest of operations, such as categorizing variables into subsets. Consider the following example of customer preferences of 1,000 respondents from California who were asked to choose between two brands of soft drinks.

	Brand A	**Brand B**	**Total**
Teenagers	400	200	600
Adults	200	200	400
Total	600	400	1,000

In this example, the respondents have been grouped into two categories: those who prefer brand A and those who prefer brand B. The respondents are further grouped based on their ages to identify the preferences of adults and teenagers for the two brands. The two subsets are mutually exclusive and exhaustive. The nominal scale in this case can be effectively used to understand the customer preferences of the respondents in two different age brackets for the two brands. However, this scale cannot be used to do advanced analysis of the responses, such as an analysis to understand how strongly the members of the two groups liked the two brands.

2. **Ordinal Scale**—An ordinal scale is used for ranking the test units or their responses to the variable under consideration. It is used for ordering only and not for comparing. For the example previously mentioned, consider four instead of two brands of soft drinks. Suppose the respondents are asked to rate each brand as one of the following: “extremely bad,” “bad,” “neutral,” “good,” or “extremely

good." The response of each respondent for all four brands will indicate his or her level of preference.

Brand	Response of a respondent
A	Good
B	Bad
C	Extremely Good
D	Extremely Bad

The ordinal scale has been used in this example to compare the four brands of soft drinks by ranking them based on the preferences of the respondents. This scale has limitations, as it cannot be used to determine the distance between the values of customer preference for different brands. So, it is not possible to find out how preferable brand C is when compared to brand A.

3. **Interval Scale**—An interval scale not only contains the characteristics of both the nominal and ordinal scale but also can be used to determine the distance between the items. The degree of difference can be expressed using this scale, but not the ratio. Temperature scales and dates are examples of interval scales.

   For the previous response, it can be inferred that the customer prefers brand C the most and brand D the least. Also, the difference between the preference for brand C and brand A can be assumed equal to the difference between the preference for brand B and brand D. Such a scale is called "assumed interval" as the difference between the different descriptors is assumed constant.

4. **Ratio Scale**—Unlike other scales, this scale deals with quantifiable data, and therefore has the concept of absolute zero. Not only does it contain the capabilities of the nominal, ordinal, and interval scales, but it also has the power of comparing the results using ratios. Actual measurements such as speed of a vehicle, age of a consumer, and number of children are all examples of ratio scales. Irrespective of the unit used to measure a particular quantity on a ratio scale, the ratio is always constant. For example, a vehicle traveling at 80 miles/hour is twice as fast as another vehicle traveling at 40 miles/hour. Even if the speed is converted to kilometers, the ratio remains constant.

Questionnaires

A questionnaire is a data collection tool that consists of a set of questions designed to obtain specific information from respondents. Correctly designing the questionnaire is a critical task because a properly designed questionnaire will enable the researcher to collect information that provides a decisive outcome. Questionnaires can be used to test hypotheses, gauge audience preferences, understand market demand for current product concepts, and decide on potential product features that may be attractive to the audience. There are a number of advantages to using a questionnaire, including the following:

- Questionnaires provide an economical and fast way of collecting data from a large audience.
- The results of a well-designed questionnaire can be easily quantified to draw actionable inferences.
- Results can be used to baseline information on various elements that can be compared over time.
- Questionnaires can be used to prove/dispel an existing hypothesis and/or create a new hypothesis.

Even though questionnaires present a quick and economical method of data collection, they are restricted in usage and have the following disadvantages:

- Questionnaires are insufficient when understanding complex issues that require in-depth research (e.g., reasons for a change in consumer perception).
- There is always a risk of interviewees not being truthful in their responses, especially for questions that are of a personal nature or ones that may be deemed embarrassing (e.g., income, how often a person goes to the restroom, etc.).
- Interviewees can misunderstand an open-ended question that is not clearly presented and this may skew the results.
- Questionnaires run the risk of being influenced by the bias of either the person framing the questions or of the person asking them.

Some of the disadvantages of a questionnaire are due to the fact that the questionnaires are open to bias introduced by the interviewer or interviewee. Some of these errors can be removed or reduced by designing the questionnaire effectively.

When designing a questionnaire, several steps are involved including identifying the objectives and hypotheses, framing the questions, selecting question formats and scales, and determining the sequence of questions.

1. **Identify objectives and hypotheses**—The key to designing the right questionnaire is to have clarity about the objectives of the survey. The target audience is identified before deciding the objectives of the questionnaire. Typically, there are two main objectives when using questionnaires to collect data:

    - to obtain relevant and accurate information from the target segment

- to test a current hypothesis or form a new hypothesis for developing a product/service

2. **Frame questions**—The most critical aspect of designing a questionnaire is to ensure that the questions are appropriately framed. For a question to be deemed appropriate for the questionnaire, it needs to fulfill the following criteria:

   - The question must be easy to understand and must not be easily misinterpreted by the respondent.
   - The answer to each question in the questionnaire should add some value to the overall survey.

3. **Decide on question formats**—Once the type of information to collect and the objectives of the research have been identified, questions that will extract relevant responses from the target segment should be crafted. Questions are typically of two types:

   - Closed-ended questions: In such questions, the respondent must choose from among the response options provided for each question.
     Example: What is your favorite season?
     a) Summer b) Autumn c) Spring d) Winter
   - Open-ended questions: These questions do not place restrictions on the answers respondents can provide. Thus, open-ended questions provide more varied responses than closed-ended ones. However, the information obtained from such questions takes greater time to analyze.
     Example: What is the most thrilling activity you have done in your life?

   In addition to crafting appropriate questions, a scale for the responses should be chosen that elicits the type of information desired by the researcher and provides valuable data.

4. **Sequence the questions**—Some general rules for sequencing questions include the following:
   - Start with a few generic questions and then move to ones asking for specific information.
   - Order questions logically from one step to another.
   - Avoid including personal and sensitive questions that would be difficult for the respondent to answer.

A few question formats should be avoided when creating a questionnaire. These include the following:
- Leading questions that are inclined to encourage the respondents to select a particular option
- Ambiguous questions that may be open to interpretation or that can be misunderstood by the respondents
- Sensitive questions that respondents may be reluctant to answer truthfully
- Double-negative questions that may be open to misinterpretation by the respondents

Example of Questionnaires:

A leading consumer electronics and home appliances company, XYZ, is conducting a survey to understand consumer buying behavior. The following are some of the questions included in its survey. In practice, when researchers design questionnaires, they try to balance the number of questions (information sought) with time taken to complete the survey. Also, depending on the quality of the information being sought, they can construct different types of questions.

Instructions: Please read each question below. Check the box next to the response that best matches your answer. Whenever a question says, "Please **check all that apply**," please check each answer that applies to you. **Your responses are completely anonymous.**

1. Have you purchased a new washing machine, refrigerator, air conditioner, TV, or any other consumer electronics or home appliances within the last year?

   - ❑ Yes
   - ❑ No → **If No, please skip to Question 17.**

2. What did you purchase? If you have purchased more than one of the products/appliances listed below during the last year, check the one that you purchased most recently.

   a. Washing Machine b. Refrigerator c. Air Conditioner d. TV

   e. Washer Dryer f. Vacuum Cleaner g. Water Purifier h. Other appliances: _______________

Please answer the following questions for **this particular appliance**.

3. Why did you purchase this new appliance? Please check only one.

   - ❑ Old one broke
   - ❑ Moved into new house
   - ❑ Remodeled
   - ❑ Needed bigger/better model
   - ❑ Didn't own one
   - ❑ Other: ____________________

4. Which company's appliance(s) did you purchase? Is it

- ❑ Brand 1
- ❑ Brand 2
- ❑ Brand 3
- ❑ Brand 4
- ❑ Brand XYZ
- ❑ Brand 6
- ❑ Brand 7
- ❑ Brand 8
- ❑ Brand 9
- ❑ Brand 10
- ❑ Other : ________________________________

5. What sources did you use to help decide which make and model to buy? Please check all that apply.

- ❑ Friends/word of mouth
- ❑ In-store materials (brochures, etc.)
- ❑ Salesperson
- ❑ Consumer reports/reviews
- ❑ Technical reviews
- ❑ Information from Internet
- ❑ Reputation of brand/manufacturer
- ❑ Other sources: ____

6. Are you happy with your purchase?

- ❑ Yes
- ❑ No → **If No, Why**? ____

7. What was important to you when deciding on the make and model? Below is a list of factors that might have influenced your decision to buy the particular model you chose. Please rate the importance of **each factor** by circling a number from 1 to 4, where 1 is extremely important and 4 is not important at all.

	Extremely Important	Important	Less Important	Not Important At All
Price	**1**	**2**	**3**	**4**
Brand	**1**	**2**	**3**	**4**
Color	**1**	**2**	**3**	**4**
Size	**1**	**2**	**3**	**4**
Premium features	**1**	**2**	**3**	**4**
Warranty	**1**	**2**	**3**	**4**

8. Did a salesperson explain the long-term costs of owning an appliance due to its energy use?

- ❑ Yes
- ❑ No
- ❑ Not applicable

9. Did you buy an energy-efficient model?

- ❑ Yes
- ❑ No
- ❑ Not applicable

10. How did you know whether or not you were buying an energy efficient model?

- ❑ Don't know
- ❑ Energy Guide label
- ❑ Consumer reports
- ❑ Sales person recommendation
- ❑ Other: ___________________________________

11. Where did you purchase the appliance?

- ❑ Locally-owned business
- ❑ Company dealer
- ❑ Company showroom
- ❑ Big-box retailer
- ❑ Online purchase
- ❑ Other: ______________________________________

12. Did you receive a Rebate/Discount on the purchase of this appliance?

- ❑ Yes
- ❑ No

If YES:

13. Was the rebate important in your decision to buy this particular make and model?

□ Yes □ No

14. What was the amount of the rebate? If you don't remember, please make your best guess. $. ________

15. Did you purchase the appliance in EMI?

- ❑ Yes
- ❑ No

16. Are you satisfied with the after sales service provided by the company?

- ❑ Yes
- ❑ No
- ❑ The product is performing satisfactorily and no after sales service has been required.

17. If you were to choose a gift from the following what would you choose? (check three)

- ❑ Cell phone (Brand: ___________________________)
- ❑ TV (Brand: ___________________________)
- ❑ Refrigerator (Brand: ___________________________)
- ❑ Washing machine (Brand: ___________________________)
- ❑ Stereo system (Brand: ___________________________)
- ❑ Air conditioner (Brand: ___________________________)
- ❑ Vacuum cleaner (Brand: ___________________________)
- ❑ Water purifier (Brand: ___________________________)
- ❑ Washer (Brand: ___________________________)
- ❑ Dryer (Brand: ___________________________)

18. In your opinion, which is the most effective mode of advertising?

a) TV b) National newspaper c) Billboard d) Magazine e) Banners and posters f) Promotional campaigns

19. In your house, who makes the majority of household decisions?

____

20. How will you rate the following companies on a scale of 1 to 7 according to their visibility (frequency of advertisement seen on TV, newspaper, billboards, etc.)?

(7 means very frequently seen and 1 means rarely or never seen)

Brand 1	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand 2	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand 3	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand 4	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand XYZ	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand 6	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Brand 7	**1**	**2**	**3**	**4**	**5**	**6**	**7**

Please answer the following questions about yourself. Your answers to these questions, as well as to all the other questions on this survey, are completely anonymous.

21. What is your age?

- Under 25
- 25–29
- 30–39
- 40–49
- 50–59
- 60+

22. Gender? a) Male b) Female

23. What is your occupation?

- Student
- Job/Salaried/Service
- Self employed
- Other: ________________________________

24. What is your annual household income (in USD)?

- Below $40,000
- $40,000–$60,000
- $60,000–$100,000
- $100,000–$200,000
- More than $200,000

25. What is your marital status?

- Single/Widowed/Divorced
- Married

26. Where do you live: ____________________________

27. What is the highest education level you have attained?

- Completed High School
- Completed College
- Completed Graduate School
- PhD/Doctorate

If you have any comments, please write them below.

Types of Quantitative Techniques

Quantitative techniques are the methods used to collect data from a larger target audience using a structured set of questions. The type of quantitative research depends on which medium can be used to reach the respondents in an optimum manner. The factors that need to be considered are the time and cost of the medium used and whether the medium obtains quality data without creating bias. The following types of quantitative techniques are available for data collection:

- **Paper and Pencil Interview (PAPI)**— The Paper and Pencil Interview (PAPI) is a quantitative technique that involves a direct interview of each respondent using a paper-based questionnaire. The use of PAPI has reduced greatly since the introduction of computers. However, it can still be convenient for non-complex questionnaires administered to small sets of people. For example, door-to-door researchers may carry copies with them, or copies may be given out to attendees at a conference or trade show.

- **Computer–assisted Personal Interview (CAPI)**—The Computer-assisted Personal Interview (CAPI) is a quantitative technique in which respondents provide their answers on a computer in the presence of an interviewer. It is considered a personal interviewing technique because the interviewer accompanies the respondent during the process. CAPI has several advantages over other forms of quantitative techniques when the sample size is large. However, this method has its own limitations as it attracts only computer-savvy respondents. This technique is mostly used when a stimulus has to be shown to the respondents and the questions are based on their understanding of the stimulus. It is often more convenient to show the stimulus on a computer screen.

- **Computer-assisted Telephonic Interview (CATI)**—The Computer-assisted Telephonic Interview (CATI) is a quantitative technique that uses computer software and a telephone system to conduct the questionnaire and record the responses. The computerized questionnaire is administered by the interviewer to the respondents over the phone. The interviewer records the responses directly into the software. This technique enables data analysis without the need for transferring responses to the computer and makes it possible for researchers to generate interim reports.

- **Computer-assisted Web Interview (CAWI)**—The Computer-assisted Web Interview (CAWI) is one of the most commonly used quantitative techniques in recent times. The questionnaires are created in software specifically built for web interviews. This interface allows for the insertion of audio-video and links to different web pages. The program changes the flow of questions depending on the respondent's answers and the available information about the respondent. It is more economical than other quantitative techniques because it does not require the interviewer to be present during the interview.

- **Central Location Test (CLT)**—The Central Location Test (CLT) is a quantitative technique in which face-to-face interviews are conducted at a centralized location such as a room or booth in a shopping mall. It is considered an effective method for testing concepts, products, packaging, and advertising effectiveness, as well as conducting sensory research. The interviewers are well trained to ensure effective data collection.

3.2.3 Outputs

3.2.3.1 Collected Primary Data*

Choosing a particular data collection method depends completely on the research problem and is influenced by the researcher's knowledge. It is essential for every researcher to identify the major factors that may influence the data collection method—cost, duration, quantity, reliability, and variety of data. The output of this process is the primary data, which can be collected through either quantitative or qualitative research, to solve a research problem. Once the data is collected, the researcher can analyze the nature of the qualitative or quantitative research and determine how that data can be applied to a given research problem's context.

The researcher should also be able to identify the limitations associated with the data and take necessary actions to ensure that the qualitative or quantitative research is as scientifically valid and reliable as possible. The primary data collected should address all major aspects of the research area—such as demographics; consumer behavior and motives; and social, economic, psychological, and personal characteristics. Choosing an inappropriate research method can impact the validity of the research problem. Therefore, it is always recommended that the research problem and objectives be understood before choosing the appropriate data collection method.

Primary Data collected could be in the following forms:

- Qualitative: transcripts, diaries, projective technique material, audio and video recordings
- Quantitative: completed questionnaires, responses from electronic surveys

Examples of Primary Data:

- GeoPro conducted a series of focus group workshops in order to strengthen the understanding of which aspects of their proposed study will be valuable to the decision-makers. The potential clients, determined through secondary research, were invited to take part in lunch-and-learn sessions hosted by GeoPro. By attending these sessions, the involved clients were given the opportunity to make suggestions and requests surrounding the type of study that GeoPro was to provide. GeoPro now had direct and valuable qualitative input about the requirements and expectations of the potential client group.
- A telecom company collects quantitative primary data from its mobile app users by prompting them to complete a customer satisfaction survey.

4. DATA PROCESSING AND DATA ANALYSIS

Once data collection is completed, the next step is to process and analyze data. Converting data into useful information that can answer the marketing research question requires processing, analysis, and interpretation of data.

It is imperative that reliable data is collected; however, even high-quality data sets can lead decision makers to make faulty decisions if the data is not processed and analyzed correctly. Therefore, it is important to have a structured framework for data processing and analysis. Decisions related to data processing and analysis should not be made after the data has been collected. While creating the research design, the researcher establishes in advance how much data will be collected and how the data will be processed and analyzed. Establishing a plan for processing and analyzing data before collection also helps ensure the following:

- that all of the required information is collected and formatted in a standardized way for easy processing;
- that unnecessary data, which will never be analyzed, is not collected; and
- that the output of the analysis will satisfy the research objectives and solve the research problem.

The transformation of raw data into actionable intelligence happens in the two phases shown in Figure 4-1.

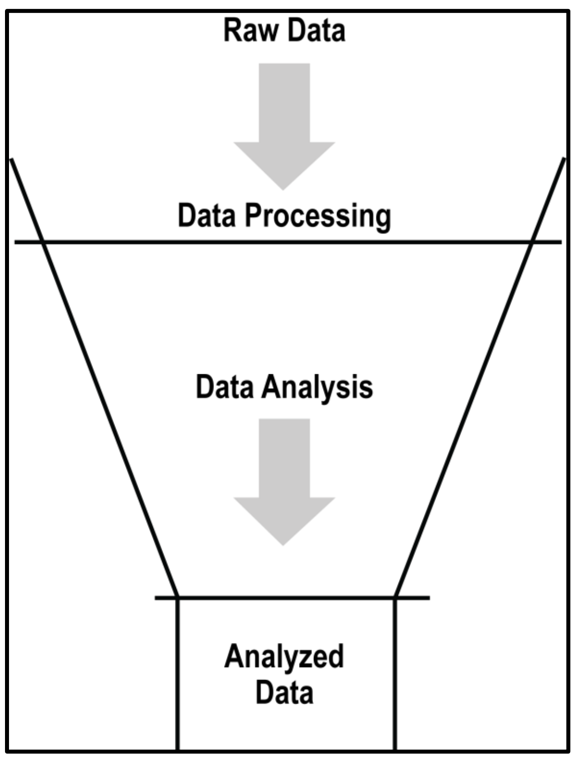


Figure 4-1: Data Transformation

It should be noted that, while a research design must be established before data collection and processing begins, researchers often produce test outputs during the collection process in order to determine the value of data being collected. Test outputs are produced in table formats and are used to make informed decisions regarding adjustments that need to be made to the research design, and any changes that are required in the data collection effort. The primary objective for the researcher is to ensure the process is yielding the data necessary for the research team to analyze the data and derive meaningful information that can be used to solve the research problem that will eventually help the marketing team make key decisions.

Figure 4-2 provides an overview of the two processes associated with Data Processing and Analysis. These are as follows:

4.1 Data Processing—In this process, various techniques and methods are used to organize, manipulate, and transform raw data into useful information ready for analysis.

4.2 Data Analysis—In this process, the application of various statistical and logical techniques and tools transforms processed data into intelligence.

4.1 Data Processing

INPUTS
1. Primary Data*
2. Secondary Data*

TOOLS
1. Coding*
2. Data Cleaning*
3. Weighting
4. Variable Respecification
5. Scale Transformation
6. Tabulation*
7. Computer Processing*
8. Data Mining*

OUTPUTS
1. Processed Data*

4.2 Data Analysis

INPUTS
1. Research Problem and Objectives*
2. Data Analysis Objectives*
3. Selected Research Design*
4. Processed Data*
5. Expert Judgment

TOOLS
1. Statistical Inference*
2. Bivariate Data Analysis
3. Multivariate Data Analysis
4. Statistical Packages*

OUTPUTS
1. Analyzed Data*

Figure 4-2: Data Processing and Analysis Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

4.1 Data Processing

After the data has been collected, it must be processed. Often, some processing is performed during data collection in order to refine continuing data collection methodologies; however, the majority of data processing is performed on completion of the data collection effort. In some cases, the raw data is in a format not suitable for data analysis. In primary data collection, raw data might be captured in an audiovisual format or in writing, which needs to be converted or formatted for processing. The raw data can also contain errors, of which there are two types—a *respondent error* occurs when a respondent makes a mistake in providing information, such as an error in a question response, and a *non-respondent error* is a mistake made by an interviewer or data collector or by a person responsible for creating an electronic data file representing the responses. The data must be cleaned and arranged in a proper format before processing can take place.

Data processing usually starts with the coding and cleaning of the data. Coding involves coding the responses, recording the codes, and more, so the collected data can be transformed into suitable tables for analysis. Data cleaning involves inspecting the data collected for omissions, errors, legibility, and consistency in classification.

The researcher has a moral obligation to present findings based on actual data that has been collected in the manner described in the research design. Even more important is the fact that the data must be error free and properly coded and cleaned so that it is useful for decision-making. If adequate attention is not paid during data preparation, the statistical results may be compromised, leading to incorrect findings and interpretations. Data integrity is an essential element for successful research and decision-making.

The primary objective of *Data Processing* is to provide processed data for data analysis. Figure 4-3 shows the inputs, tools, and outputs for the *Data Processing* process.

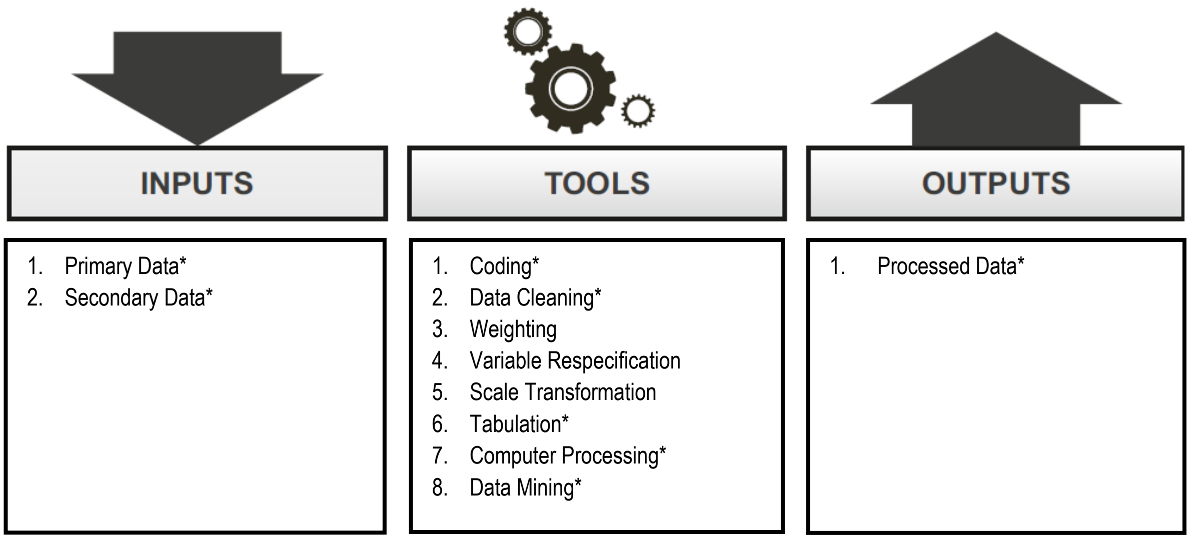


Figure 4-3: Data Processing—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

4.1.1 Inputs

4.1.1.1 Primary Data*

When a marketing research project involves the collection of primary data, data processing is a critical component of the data collection effort. When first collected, raw data are usually in no condition for interpretation. The raw data will contain respondent-level data that will only be useful after the data has been converted or reformatted so that it can be properly analyzed. Also, the possibility of various data collection errors is high when data is collected for the first time. The raw data may contain errors, omissions, and inconsistencies. Errors can be categorized as the following:

- **Fieldworker Error or Non-respondent Error**: These errors are made by the fieldworkers or administrators who administer the primary data collection by conducting interviews, surveys, or focus group discussions. Even with the use of professional data collection personnel, it is impossible to completely avoid this type of error. The errors may be intentional or unintentional. Such errors might include the incorrect recording of a subject's demographic information or a typographical error during data entry.

- **Respondent Errors**: These errors occur as a result of something the respondent does or does not do and may occur irrespective of data collection methods. Respondent errors might also be intentional or unintentional. Such errors may be the result of misinterpreting a question or incorrectly completing a form.

Apart from collating the unstructured data from various individual forms and organizing the data into a structured format, the researcher also needs to focus on cleaning the data by removing any errors. The raw primary data need to be validated, coded, and stored electronically.

Example of Primary Data:

- Restaurants often solicit feedback from customers in the form of a questionnaire. Once sufficient feedback forms are collected, they are analyzed to gain insights on various aspects of service. While processing the feedback forms, extra care needs to be taken to clean the data by removing any errors. Often, departing customers provide hurried feedback, putting no thought into their efforts. They tend to choose random options leading to errors. Also, if the feedback is gathered by the wait-staff, the servers might tweak the responses in unsealed questionnaires to get favorable outcomes for themselves. While processing the feedback forms, the data analyzer should consider these possible errors. If the responses indicate obvious inaccuracy, they may be dropped. Reviewers need to make decisions based on their own judgment.

4.1.1.2 Secondary Data*

When secondary data collection is involved in a research project, the data processing activity may not be as extensive as in the case of primary data. In many cases, the use of secondary data eliminates the data processing activity completely because the data had already been processed when the original study was conducted. However, the researcher should always keep in mind that the original data might contain errors—it is therefore inadvisable to base any major decision exclusively on a single source of secondary data. If possible, multiple sources of secondary data should be considered for cross-checking. Also, the secondary data may be outdated or less relevant to the current project. The researcher may request the publishing organization to furnish information on the collection instruments and methods. Good documentation regarding secondary data provides necessary background and much-needed context, making the re-use of data worthwhile and systematic. It is important to contextualize and reconstruct the secondary data for the current project. If there is any difficulty in verifying the accuracy of secondary data, the researcher should proceed with extreme caution. Researchers naturally prefer data from reliable sources, such as government agencies or reputable organizations. When the format of the secondary data is different than the expected format for the current project, data conversion or transformation may be necessary. Data may need to be altered from the original form to a format more suitable for achieving a stated research objective.

Regardless of whether the secondary data have been verified, are in compatible form, or require modifications, the researcher must always be careful to not introduce errors into data.

4.1.2 Tools

4.1.2.1 Coding*

Coding is the first task in *Data Processing*. Data coding is the process of organizing, labeling, and sorting the data into appropriate headings to make it convenient for data analysis in both quantitative and qualitative research. In data coding, responses are usually assigned numbers and symbols to make them measurable and recordable. Categories and character symbols for responses must be established before tabulation. The numbers and symbols for categorizing the data are called codes. Codes allow the researcher to reduce large quantities of information into a form that can be easily processed and analyzed, particularly by computer based statistical tools.

Coding starts with specifying the different categories or classes into which the responses are to be classified and ends with allocating codes to individual answers. The process of creating codes can be both preset and open. Some codes are predefined by the researcher. It is beneficial to start with a list of preset codes derived from the conceptual framework, the list of research questions, the problem areas, and other significant identifiers. Prior knowledge of the subject matter helps the researcher create these codes. While it is good to begin data collection and coding with preset codes, another set of codes will emerge or can be developed from reading and analyzing the data. These “emergent codes” are those ideas, concepts, actions, relationships, and meanings that appear in the data and that are different than the preset codes. The researcher tries to integrate predefined categories in the data collection forms using closed questions so that it becomes easy to transfer the data into a tabulated format later. The use of precoded closed questions reduces the cost of coding and also improves accuracy. Sometimes this is not feasible, and the researcher needs to ask open-ended questions in the survey. Including open-ended questions in a survey garners qualitative data, as opposed to the quantitative data that results from multiple-choice questions. The coding for the open-ended questions is performed after collection. Once coded, these answers can be analyzed in the same way that multiple-choice, closed questions are analyzed.

Coding is an evolving process when it comes to qualitative studies. As data is coded, the coding categories can be refined further. Researchers may add, collapse, expand, and revise the coding categories. In some cases the codes may be too broad, and subcategories may need to be added. In other cases the codes designed may be too specific and capture all possible responses. In these cases, the researcher must consider how to combine categories into a broader idea. When coding, the researcher should always make the codes fit the data, rather than trying to make the data fit the codes.

Example of Coding:

- A question regarding the purchase frequency of a respondent can be precoded using various numbers.

Question	***Answers***	***Codes***
How often do you purchase the product?	Never	*1*
	Annually	*2*
	Monthly	*3*
	Weekly	*4*
	Daily	*5*

- A question regarding job satisfaction of a respondent can also be precoded.

Question	***Answers***	***Codes***
How satisfied are you in your job?	Very Dissatisfied	*1*
	Dissatisfied	*2*
	Neutral	*3*
	Satisfied	*4*
	Very Satisfied	*5*

4.1.2.2 Data Cleaning*

Data cleaning is the step in data processing that follows coding. After the data have been coded, the data must be cleaned and checked for outliers before proceeding to analysis. Fieldwork often produces data containing errors. Data cleaning is the process of checking the data for omissions, consistency, and legibility. This step can involve, for example, checking for errors and omissions on questionnaires or other data collection forms. When researchers discover a problem or inconsistency, they need to make necessary modifications in order to ensure the data is readable. If a particular data point is beyond comprehension or cannot be corrected (by re-administration), it needs to be excluded from the analysis.

When raw data is collected for the first time, the researcher needs to examine the data points for inconsistencies and determine if an inconsistency is an outlier and should be discarded, or whether the entire questionnaire should be discarded and/or re-administered. The researcher needs to follow established guidelines of data and quality checks to determine which answers are inconsistent and choose an appropriate course of action. Sometimes, the data set or responses may contain only a partial or a vague response. If possible, the researcher should contact the respondent to gain clarity. In some situations, the predefined protocol is to not pursue the missing data and simply leave the question blank. In the case of multiple or duplicate markings on a single question, researchers must carefully examine the answers and keep the most accurate one based on their judgment. If no clear interpretation can be made, the response should be discarded. If a respondent provides a long statement, the researcher may need to decipher the underlying meaning to see if the answer is appropriate for that category. In some instances, researchers may decide to change the category code based on their judgment. This also applies to any out-of-place responses.

Two types of data cleaning methods are used: field cleaning and central or in-house quality checks. The use of each type is based on the time when the data cleaning is conducted. Fieldworkers are often responsible for conducting preliminary field data cleaning and quality checks on the same day as the interview, usually because field interviewers often use their own abbreviations and shorthand while recording responses. To prepare the data for processing, they must complete the data collection by replacing their shorthand with the full responses before the data can be forwarded for processing. Field supervisors are also often responsible for identifying technical omissions such as a blank page on an interview form, checking the legibility of handwriting on open-ended responses, and clarifying responses that are logically or conceptually inconsistent. A daily field quality check may also allow fieldworkers to identify respondents who should be contacted again in a timely fashion to address any omissions.

After questionnaires are collected by the researchers, central or in-house data cleaning is performed. In-house data cleaning involves rigorously investigating the results of the data collection. The research supplier or research department typically has centralized office staff who perform the coding and data cleaning function.

Ideally, a single researcher should carry out the entire data cleaning task, so that consistency is ensured throughout the process and across data sets. In large scale projects, however, this often is not feasible. A systematic procedure for assessing the questionnaires should be developed by the researcher, so that others have clearly defined decision rules to follow while checking the data. Any inferences—such as missing values—should be made in a manner that limits the chance for the researcher's subjectivity to influence the response. When multiple researchers are involved, uniform guidelines should be developed so that maximum possible consistency can be achieved.

While it is the job of the field supervisor and researcher to check the entire data set for completeness, accuracy, and consistency, the researcher should also ensure that not too much time is spent on trivial or relatively minor errors.

Example of Data Cleaning:

- All primary research can suffer from several types of errors in data collection. One type of error is incorrect reporting or recording of the characteristics of individuals' experiences with a product. This type of error is usually a fieldworker mistake, such as mistranslating, omitting questions, incorrectly recording responses, or misunderstanding a respondent when completing a behavioral survey. Data cleaning must be conducted to ensure the validity and consistency of individual records and relationships among records collected in a single store location and by the same fieldworker to check the reasonableness of the aggregated data. Another type of error is highly unlikely or impossible data records, such as a three-year-old child documented with a Ph.D. or a male who reported giving birth in the past year. Data cleaning reduces distorted estimates, facilitates processing, and increases user confidence in the data and its usability in managerial decision-making.

4.1.2.3 Weighting

Weighting is a tool used to assign each respondent in the database a value or multiplier that accounts for the extent to which certain categories of respondents should contribute to the overall results. Some categories, for example, may be considered more significant in the overall results than others. To make the analysis more inclusive, a higher weightage can be given to the categories with fewer respondents while a lower weightage can be given to the categories with abundant responses. Weighting is mostly used to make the sample data more representative of a target population. Because every database has varying numbers of responses from different categories of respondents, it is essential that higher weightage is given to the respondents with the desired characteristics.

Example of Weighting:

- If a researcher wants to understand the consumer preference for a new range of luxury watches, he might give higher weightage to older and experienced respondents. Thus, the researcher might want to assign a 1.0 weightage to the respondents below the age of 25 years, 2.0 to the respondents in the age group of 25–39 years, and 3.0 to the age group of 40 years and older.

4.1.2.4 Variable Respecification

The main purpose of respecification is to transform the data to create new variables, or to modify existing variables, and ensure that these variables are more consistent with the research objective. For example, a variable measuring a respondent's average consumption of a product per year could be transformed from gross numbers into categories: consumption below 100 Liters is Low, 101 to 250 Liters is Medium, and over 500 Liters is High. Thus, a simple modification to an existing variable can create a new variable. Dummy variables are also commonly used to specify a particular set of data. The most commonly used dummy variables are binary variables. For example, if a researcher is conducting research on wine consumption, before exploring the transformation of consumption into low, medium, and high for qualifying the drinkers, he might run a binary variable on the data set to remove all non-drinkers. By using 0 to indicate a non-drinker and 1 to indicate a drinker in the marketing research, the researcher has created a mutually exclusive category with the 0 that will remove all non-drinkers from further analysis.

4.1.2.5 Scale Transformation

In many instances, collected data must be compared with another data set. In order to make the data suitable for comparison, the scale needs to be transformed to ensure comparability with the scale values of the reference data set. The process of transforming the data to a meaningful, comparable format is known as scale transformation. There are a variety of appropriate scales to choose from, based on the specific requirements of the research project. The seven-point semantic differential scale, the continuous rating scale, Stapel's scale, multidimensional scaling, and the five- or three-point Likert scale are several scales that can be used to measure variables such as image, attitude, and lifestyle. In order to compare image scores with attitude scores or lifestyle scores, it is necessary to transform the scales used to measure these variables. In some cases, the same scale is used for different variables: some responses use the higher end while others use the lower end of the rating scale. Scale transformation is used to correct and appropriately transform the data.

The following are the most commonly used scales:

- **Seven-point semantic differential scale**—The semantic differential scale is a scale used to discover respondents' attitudes toward a product or service. This scale asks respondents to rate a product, service, brand, or company based on a seven-point rating scale that has two bipolar adjectives at either end. It is designed to measure the connotative meaning of objects, events, and concepts. This scale is useful in helping compare qualitative language that a marketer is using, or plans to use, in advertising and to create brand associations for the material and value proposition.

Example of seven-point semantic differential scale:

- A marketer seeking customer ratings regarding guests' hotel experiences might ask the guests to rate their stay at the hotel by placing an X (or circling) the term that best represents how the customer feels about his or her stay. The questionnaire using the 7-point semantic differential scale might appear as follows:

  Memorable _:_:_:X:_:_:_ Forgettable
  Excellent_:_:X:_:_:_:_Awful
  Luxurious_:_:X:_:_:_:_Cheap
  Exciting_:_:X:_:_:_:_Boring
  Safe_:_:X:_:_:_:_Dangerous

  By identifying certain clientele (i.e., demographic/psychographic segments) marketers can test language that reflects how the business is positioning its brand to gauge not only service or experience measures, but also to test different customer experience feedback cards using different pairs of words to gauge impact on ratings.

- **Continuous rating scale**—The continuous rating scale, also known as the graphic rating scale, is used where the respondents are asked to mark a point on a scale that they find appropriate. They are used to measure categories such as satisfaction, brand awareness, and quality in relation to a particular product or service. A continuous rating scale simplifies the response decision and can be used in cases where more complex language skills are not appropriate, such as when children or individuals with a language barrier are making the ratings.

- **Five-point Likert scale**—A Likert item is a statement that a respondent is asked to evaluate. A Likert scale is the sum of responses to several Likert items. The Likert scale is useful for evaluating a respondent's purchasing behavior, opinion of a product, or level of product satisfaction.

Used as an ordinal scale measure of attitudes, beliefs, and opinions, Likert scales are used frequently and are easily understood by respondents. The Likert scale does not force a response and can accommodate a degree or range of acceptance.

4.1.2.6 Tabulation*

Tabulation is one of the most frequently used tools in data analysis. It is used to organize the data into predefined categories and to count the number of observations from the data set that occurs in each of the predefined categories. When only one variable is involved in the analysis, it is called univariate tabulation. When two or more variables are involved, it is called bivariate or multivariate tabulation, respectively. Tabulation is a useful tool for data cleaning, for finding out the frequency distribution of the variable, and for calculating the descriptive statistics of the variable in question. Tabulation can be performed both manually and by machine, often computers. Manual tabulation is typically done when the variables and categories involved are few, the sample size is small, and the analysis required is basic. For large data sets with multiple variables, machine tabulation is more efficient.

Tabulation in which one variable is involved is simple to organize and analyze. However, in many cases, responses for two or more questions need to be tabulated together. This requires a different technique called "cross-tabulation," also known as cross-tabs and contingency table analysis.

Tabulation helps in various types of data analysis needs. While frequency distribution is the most basic output from tabulation, it can also be used to calculate descriptive statistics such as mean, median, and mode. Some of the basic statistics associated with frequency distribution are as follows:

- Measures of central tendency
- Measures of dispersion
- Measures of symmetry

Please refer to Appendix A.3.2 for a detailed discussion of descriptive statistics.

Example of Tabulation:

- A question regarding the purchase frequency of respondents can be tabulated as follows:

Question	***Answers***	***Number of respondents***	***Percentage of respondents***
How often do you purchase the product?	Never	*20*	*40*
	Annually	*10*	*20*
	Monthly	*5*	*10*
	Weekly	*10*	*20*
	Daily	*5*	*10*

4.1.2.7 Computer Processing*

The collected data must be edited, coded, and tabulated for effective data processing and analysis. Data can be handled manually, but this is time-consuming and the possibility of manual errors is significant. Computer processing uses computer software to summarize, analyze, and convert data into usable information. Computers help in terms of accuracy, efficiency, and sophistication. For large amounts of data or surveys, computer processing becomes both convenient and economical. For such large surveys, the questionnaires are coded into the computer in advance, thus simplifying and improving efficiency in the later stages of the process. There are several computer software applications available to aid data processing and analysis..

It is essential for the researcher to know the relationship between data collection, preparation, processing, and analysis. Once the data is entered into the computer, the researcher can carry out many checks and cross-data analyses in the context of the research problem.

The major outputs of data that have been computer processed are in the form of tables, charts, and graphs. The tables help identify the number of respondents in each segment or category of answers and their percentage.

Example of Computer Processing:

- SPSS, SAS, and MATLAB are some of the popular computer programs used for data processing and analysis.

4.1.2.8 Data Mining*

Data mining is the most used secondary data processing tool. Through data mining, researchers can discover unknown valid information from huge databases; these discoveries help organizations make key business decisions. In other words, data mining is an exploratory data analysis without any prior hypothesis. Data mining is also an automated process of data analysis where the software or the system fetches valid information from the database. Data mining helps organizations understand consumer buying behavior and trends for the near future; this process can also predict changes. Data mining is an important tool that allows researchers and analysts to manage businesses effectively and make key strategic decisions. Data mining is relatively easy when organizations have their own data warehouse. However, it is not necessary to have a data warehouse to perform data mining.

Data mining operations are classified into four major categories: predictive modeling, database segmentation, link analysis, and deviation detection.

- **Predictive Modeling**—This is a form of inductive reasoning that takes specific data or information and makes a broader generalization. This technique uses neural networks and inductive reasoning algorithms to predict future trends. The conclusion from this method is always predictive and may not be accurate.
- **Database Segmentation**—This is used to classify the data into clusters. The segmentation is done using statistical cluster analysis techniques.
- **Link Analysis**—This approach involves determining the associations between data and sequential patterns. This method helps discover the association between two or more data records, which in turn helps determine a possible solution to the research problem.

- **Deviation Detection**—This method helps researchers determine the data or records that cannot be considered or included in the analysis. In this method data are identified and removed if they are not within the scope of the research problem.

Examples of Data Mining:

- E-commerce websites use data mining in online marketing efforts. A customer creates an account and makes a purchase; this purchase information is stored in the company database. When customers return to the website, the website presents them with some recommendations based on their previous purchases. For instance, if a customer purchases a camera, on the second visit, the website will recommend lenses and camera bags. The company may also send e-mails recommending appropriate products based on purchase history.
- Online tourism portals offer worldwide holiday packages. Data mining software can analyze customers' historic web analytics data based on the customers' clicks and website actions. This in turn enables the tourism portals to differentiate the customer segment and recommend appropriate packages to them.

4.1.3 Outputs

4.1.3.1 Processed Data*

Processed data is the output of this process. The data, once processed, is ready for data analysis. It has been refined and formatted from its original raw data state and is in a state that enables the researcher to efficiently and effectively analyze the results. In some cases, Content Analysis Grids are developed in order to see the data in a categorized manner. Processing the data involves systematic procedures such that the outputs are easily accessible for analysis. The processed data should be in a format that facilitates the necessary analysis to answer the research problem and meets the research objectives completely.

4.2 Data Analysis

Once the *Data Processing* process is complete, the data is ready to be analyzed using various tools and techniques that will help the researcher discover useful information, which in turn will suggest conclusions and help in decision-making. *Data Analysis* involves summarizing the collected data and organizing the data in such a manner that the researcher can solve the research problem and meet the research objectives. In other words, *Data Analysis* transforms data into intelligence.

Figure 4-4 shows the Inputs, Tools, and Outputs associated with this process.

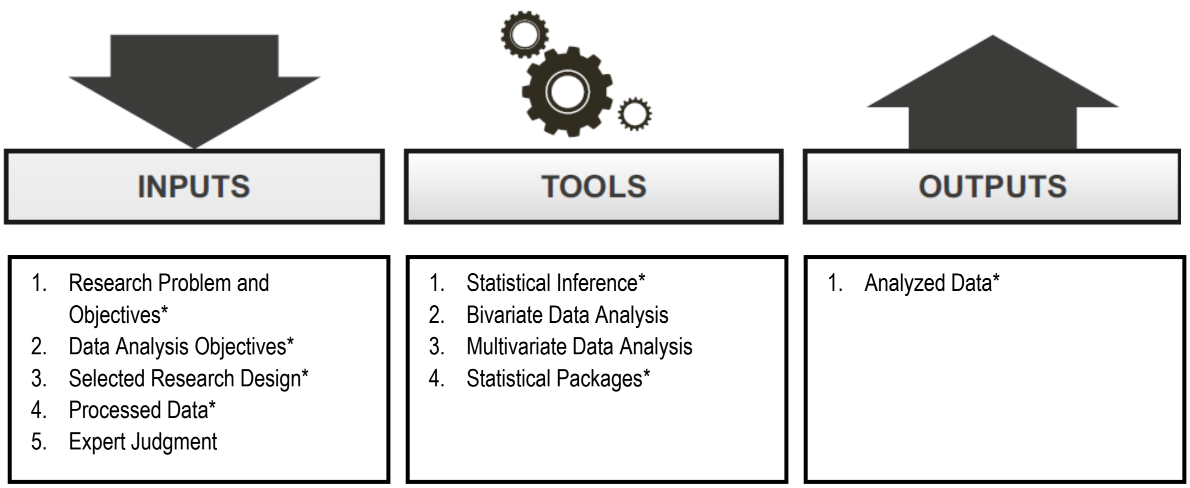


Figure 4-4: Data Analysis—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

4.2.1 Inputs

4.2.1.1 Research Problem and Objectives*

Understanding the research problem and objectives is a key component of data analysis. This information helps the researcher establish objectives with regards to data analysis, choose an appropriate analytic approach, and define appropriate boundaries for analysis. An effective analysis of the data is impossible without knowing what ultimately must be achieved, and knowing the research problem and objectives ensures focus throughout the data analysis process. Conducting analyses without a clear idea of the objectives can result in wasted resources and, sometimes, inadequate analyses.

4.2.1.2 Data Analysis Objectives*

Among the data analysis objectives are two key components: ensuring the results are derived from high quality data and effectively analyzing the data in order to answer the research question. Identifying achievable data analysis objectives involves having a clear understanding of the relationship between the data and the analysis objectives, knowing how to best approach and assess the research problem, ensuring the quality of data available for collection meets expected standards, and knowing the resources and time available. It is important to have a comprehensive understanding of these factors before commencing data analysis. The success of the research project will require a focused approach to both meeting the data analysis objectives and ultimately solving the research problem.

4.2.1.3 Selected Research Design*

The selected research design is an output of the *Choose the Research Design* process discussed in section 2.2. The selected research design is a step-by-step plan that guides data collection and analysis. An effective analysis should enhance the overall quality of data and improve the overall data collection process by determining whether the data collection needs to be expanded, reduced, or refined. Data analysis techniques may vary and can be flexible depending on the selected research design. The research design leads the researcher to use the best-suited model for the analysis. Additionally, the selected research design dictates the types of variables to be used in a questionnaire or study, which in turn determines the most effective analytical model for getting the desired results from the study. The research design guides the researcher to understand the type of data analysis that can be done in the study. The researcher is not limited to any technique when analyzing a particular data set. However, there are many factors that need to be considered to ensure that the analysis is appropriate and efficient and that the output is meaningful.

Example of Selected Research Design:

- In a research project, to understand the impact of certain factors (e.g., in-store promotions, temperature, holidays, and ad campaigns) on the sales of a particular item, the researcher can use various types of predictive techniques and analytical models. In this instance, a causal research design would be established that would control certain variables in order to determine specific relationships between certain variables and customers' purchasing behavior. Using various data analysis approaches, the researcher can determine, from the output of the analysis, the likelihood of increased sales based on changes to certain variables, such as discounts and promotions on the item.

4.2.1.4 Processed Data*

The processed data is an output of the *Data Processing* process described in section 4.1. Processed data is data that has been refined and formatted from its original raw data state. Processed data is an important input in data analysis because performing analysis on raw data is extremely cumbersome and sometimes impossible. An effective data processing step will result in quality data in the right format to be analyzed, such that the analysis outcomes answer the research problem and meet the research objectives. The quality of the output of data analysis is greatly affected by the data used to perform statistical techniques.

Example of Processed Data:

- A food oil manufacturer was trying to understand its product positioning effectiveness. The research team identified the derived importance of product parameters based on a comparative rating of other edible oil brands on the same set of parameters (e.g., cost of oil, p-Anisidine value, fatty acids content, fatty acid profile, and iodine value). The primary positioning was on lightness and low cholesterol content. However, the processed data highlighted that the user ratings were more skewed toward the cost of the oil. Thus, the processed data revealed the core driver of the category—pricing/cost of oil—which the marketing team needed to know in order to understand the effectiveness of the product's positioning.

4.2.1.5 Expert Judgment

Researchers often seek expert judgment for complex data analysis. The knowledge and experience of an expert can provide a focused approach to analyzing specific objectives and solutions.

Analyzing previously collected data can provide the researcher with enough information for the marketing research project. However, some research problems do not have ample historical data available or others require huge amounts of capital. The researcher may not be able to conclude the research with qualified data. In such instances, the researcher uses the knowledge and experience of an expert to seek guidance for analysis. Furthermore, consulting subject matter experts and benefitting from their experience and knowledge greatly mitigates the risk of failure in any given project.

Example of Expert Judgment:

- A leading, European-based manufacturer of precision instruments wanted to establish its presence in a new market: India. It required the details of various established dealers in India that dealt with products that were similar to those they manufactured. The requirement was in the domain of dealer mapping and the research assignment was handed over to a marketing research organization in India. The research team needed to collect the contact information of about 5,000 dealers across India. The European firm wanted the vast amount of information mined and collated within a very short timeframe, and the research team did not have enough time to collect that much primary data by the deadline. However, having operated in India for several years and having done similar projects in the domain, the research team managed to collect secondary data from trustworthy sources. The team also had access to subject matter experts within the company who analyzed the processed data quickly. With their experience in the Indian market and information sources, they identified the right data to analyze without wasting time on irrelevant data. This is something that the manufacturer's internal team could not have done in the short timeframe.

4.2.2 Tools

4.2.2.1 Statistical Inference*

Statistical inference provides interpretations of and about a population from a sample. In practice, it is often impossible to analyze the entire data of a population due to constraints such as time, cost, and manpower; when this is the case, a sample is selected that is treated as representative of the population, and further statistical analysis is performed to interpret the population's characteristics.

Two topics of discussion are important in statistical inference: Estimation and Hypothesis Testing.

1. **Estimation**

   In real-life problems, the true values of each of a population's parameters (e.g., population mean and variance) are not known, and statisticians usually determine an estimated or approximated value using samples drawn from the population. In the process of determining an estimated value of a population parameter, an error is expected to arise. The theory of estimation is all about minimizing the error and establishing a measure of accuracy when estimating population parameters. Estimation has two broad classifications: Point Estimation and Interval Estimation.

- **Point Estimation**

  When a random sample is drawn from a population, population parameters can be estimated using sample statistics. The estimated value determined is the point estimate of the associated population parameter. The disadvantage here is that one does not know the level of accuracy for the estimate obtained.

Example of Point Estimate:

- Suppose an estimate is required for the average age of a sample of 1,200 people who use a particular brand of soap. Instead of reviewing all 1,200 survey responses, the researcher can randomly select a sample of 50 for measurement. The average and standard deviation are 40.1 and 0.006, respectively. This average value, 40.1, is called the point estimate.

In this case, the sample mean (statistic) is an estimator of the population mean (parameter). A statistic is a value obtained from a sample, while a parameter is a value from the population.

The standard deviation of the distribution of means indicates the standard error (SE), which is the amount of error that can occur when a sample mean is used for estimating a population mean. See Appendix A.3.5.1, “Standard Error,” for more details.

- **Interval Estimation**

  When an interval or range is determined within which the true value of the population parameter lies, there is a measure of accuracy or level of confidence that the estimated value of the population parameter falls within the interval. This interval is known as the confidence interval (CI) for the associated population parameter. The level of confidence is expressed as a percentage.

Typical Symbols Used in Estimation

	Sample	Population
Mean	$\overline{x}$	μ
Number of values (size)	n	N
Standard deviation	s	σ

4

Construction of Confidence Interval (CI)

The formula for the endpoints of the CI for population mean μ is: $\left(\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$ when population standard deviation σ is known. (See appendix A.3.5.4 for information on how this formula has been derived.) When σ is unknown, it can be estimated by sample standard deviation s.

α = probability that the population mean does not fall in the interval (α -risk).

$1 - \alpha$ = probability that the population mean is in the interval (confidence level).

$z_{\frac{\alpha}{2}}$ = the value from the Z-table (standard normal table) with an area of α/2 to its right.

Example of Interval Estimation:

- Consider the previous example about the average age of a soap user. How do researchers know that the population mean is exactly 40.1?

  They probably cannot know, due to sampling error; that is, if some other sample is taken, the average may not be the same as 40.1. In other words, researchers may get different point estimates every time they pick a random sample of the same size from the population.

  To mitigate this problem, a technique is required to determine how accurate the point estimate is. That technique is known as Interval Estimation.

  In this problem, $\overline{x}$ = 40.1, n = 50, N = 1200, 1 – α = 95% = 0.95, α = 0.05, and s = 0.006.

  (See appendix A.3, "Statistics," to learn how to calculate mean and standard deviation.)

  Formulas for the endpoints of the CI are: $\left(\overline{x} - z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right), \overline{x} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)\right)$

  where $Z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$% of the standard normal distribution.

  From the z-table, the z-value with 2.5% to its right is 1.96 (i.e., $Z_{\frac{\alpha}{2}}$ = 1.96).

  Calculation of the end points of the confidence interval:

  $\overline{x} + Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$ = 40.1 + 1.96 * $\left(\frac{0.006}{\sqrt{50}}\right)$ = 40.1017

  $\overline{x} + Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$ = 40.1 - 1.96 * $\left(\frac{0.006}{\sqrt{50}}\right)$ = 40.0983

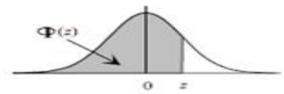
  Use Excel to calculate the following:

  =40.1+1.96*(0.006/SQRT(50))

  =40.1-1.96*(0.006/SQRT(50))

  Therefore, we are 95% confident that the population mean μ is between 40.0983 and 40.1017.

  Thus, it can be stated, "We are 95% confident that the population mean is between 40.0983 and 40.1017" or equivalently "The 95% confidence interval for the population mean is (40.0983, 40.1017)."



z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	Φ(z)	z	
1.20	0.88493	1.60	0.94520	2.00	0.97725	2.40	0.99180	2.80	0.99744	3.20	0.99931	3.60	0
1.21	0.88686	1.61	0.94630	2.01	0.97778	2.41	0.99202	2.81	0.99752	3.21	0.99934	3.61	0
1.22	0.88877	1.62	0.94738	2.02	0.97831	2.42	0.99224	2.82	0.99760	3.22	0.99936	3.62	0
1.23	0.89065	1.63	0.94845	2.03	0.97882	2.43	0.99245	2.83	0.99767	3.23	0.99938	3.63	0
1.24	0.89251	1.64	0.94950	2.04	0.97932	2.44	0.99266	2.84	0.99774	3.24	0.99940	3.64	0
1.25	0.89435	1.65	0.95053	2.05	0.97982	2.45	0.99286	2.85	0.99781	3.25	0.99942	3.65	0
1.26	0.89617	1.66	0.95154	2.06	0.98030	2.46	0.99305	2.86	0.99788	3.26	0.99944	3.66	0
1.27	0.89796	1.67	0.95254	2.07	0.98077	2.47	0.99324	2.87	0.99795	3.27	0.99946	3.67	0
1.28	0.89973	1.68	0.95352	2.08	0.98124	2.48	0.99343	2.88	0.99801	3.28	0.99948	3.68	0
1.29	0.90147	1.69	0.95449	2.09	0.98169	2.49	0.99361	2.89	0.99807	3.29	0.99950	3.69	0
1.30	0.90320	1.70	0.95543	2.10	0.98214	2.50	0.99379	2.90	0.99813	3.30	0.99952	3.70	0
1.31	0.90490	1.71	0.95637	2.11	0.98257	2.51	0.99396	2.91	0.99819	3.31	0.99953	3.71	0
1.32	0.90658	1.72	0.95728	2.12	0.98300	2.52	0.99413	2.92	0.99825	3.32	0.99955	3.72	0
1.33	0.90824	1.73	0.95818	2.13	0.98341	2.53	0.99430	2.93	0.99831	3.33	0.99957	3.73	0
1.34	0.90988	1.74	0.95907	2.14	0.98382	2.54	0.99446	2.94	0.99836	3.34	0.99958	3.74	0
1.35	0.91149	1.75	0.95994	2.15	0.98422	2.55	0.99461	2.95	0.99841	3.35	0.99960	3.75	0
1.36	0.91308	1.76	0.96080	2.16	0.98461	2.56	0.99477	2.96	0.99846	3.36	0.99961	3.76	0
1.37	0.91466	1.77	0.96164	2.17	0.98500	2.57	0.99492	2.97	0.99851	3.37	0.99962	3.77	0
1.38	0.91621	1.78	0.96246	2.18	0.98537	2.58	0.99506	2.98	0.99856	3.38	0.99964	3.78	0
1.39	0.91774	1.79	0.96327	2.19	0.98574	2.59	0.99520	2.99	0.99861	3.39	0.99965	3.79	0
1.40	0.91924	1.80	0.96407	2.20	0.98610	2.60	0.99534	3.00	0.99865	3.40	0.99966	3.80	0
1.41	0.92073	1.81	0.96485	2.21	0.98645	2.61	0.99547	3.01	0.99869	3.41	0.99968	3.81	0
1.42	0.92220	1.82	0.96562	2.22	0.98679	2.62	0.99560	3.02	0.99874	3.42	0.99969	3.82	0
1.43	0.92364	1.83	0.96638	2.23	0.98713	2.63	0.99573	3.03	0.99878	3.43	0.99970	3.83	0
1.44	0.92507	1.84	0.96712	2.24	0.98745	2.64	0.99585	3.04	0.99882	3.44	0.99971	3.84	0
1.45	0.92647	1.85	0.96784	2.25	0.98778	2.65	0.99598	3.05	0.99886	3.45	0.99972	3.85	0
1.46	0.92785	1.86	0.96856	2.26	0.98809	2.66	0.99609	3.06	0.99889	3.46	0.99973	3.86	0
1.47	0.92922	1.87	0.96926	2.27	0.98840	2.67	0.99621	3.07	0.99893	3.47	0.99974	3.87	0
1.48	0.93056	1.88	0.96995	2.28	0.98870	2.68	0.99632	3.08	0.99896	3.48	0.99975	3.88	0
1.49	0.93189	1.89	0.97062	2.29	0.98899	2.69	0.99643	3.09	0.99900	3.49	0.99976	3.89	0
1.50	0.93319	1.90	0.97128	2.30	0.98928	2.70	0.99653	3.10	0.99903	3.50	0.99977	3.90	0
1.51	0.93448	1.91	0.97193	2.31	0.98956	2.71	0.99664	3.11	0.99906	3.51	0.99978	3.91	0
1.52	0.93574	1.92	0.97257	2.32	0.98983	2.72	0.99674	3.12	0.99910	3.52	0.99978	3.92	0
1.53	0.93699	1.93	0.97320	2.33	0.99010	2.73	0.99683	3.13	0.99913	3.53	0.99979	3.93	0
1.54	0.93822	1.94	0.97381	2.34	0.99036	2.74	0.99693	3.14	0.99916	3.54	0.99980	3.94	0
1.55	0.93943	1.95	0.97441	2.35	0.99061	2.75	0.99702	3.15	0.99918	3.55	0.99981	3.95	0
1.56	0.94062	1.96	0.97500	2.36	0.99086	2.76	0.99711	3.16	0.99921	3.56	0.99981	3.96	0
1.57	0.94179	1.97	0.97558	2.37	0.99111	2.77	0.99720	3.17	0.99924	3.57	0.99982	3.97	0
1.58	0.94295	1.98	0.97615	2.38	0.99134	2.78	0.99728	3.18	0.99926	3.58	0.99983	3.98	0
1.59	0.94408	1.99	0.97670	2.39	0.99158	2.79	0.99736	3.19	0.99929	3.59	0.99983	3.99	0
1.60	0.94520	2.00	0.97725	2.40	0.99180	2.80	0.99744	3.20	0.99931	3.60	0.99984	4.00	0

Standard Normal Table

The preceding table represents the percentage of the shaded region, denoted as $\Phi(z)$, under the standard normal curve corresponding to the z value. Therefore, $Z_{\frac{\alpha}{2}}$ is the value of z corresponding to $\Phi(z) = 0.975$ because $Z_{\frac{\alpha}{2}}$ is the upper 2.5% point of the standard normal distribution. In other words, it is the value of the standard normal variable z which covers 97.5% of the area on its left and 100 – 97.5 = 2.5% of the area on its right.

Confidence Interval Conditions

When the standard deviation of a population is unknown, the sample standard deviation s can be used as an estimate of the population standard deviation σ.

Equations for determining confidence interval are given for various conditions:

Population Parameter	Conditions	CI Formula
Population Mean (μ)	Large sample (n ≥ 30) when σ known	$\overline{x} - Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$
Population Mean (μ)	Large sample (n ≥ 30) when σ unknown	$\overline{x} - Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + Z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)$
Population Mean (μ)	Small sample (n < 30) when σ unknown and population is Normally distributed	$\overline{x} - t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right)$
Population Standard Deviation (σ)	Population is Normally distributed	$\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right),n-1}}}$

Note that it is not necessary to remember the formulas for confidence interval because all of these can be obtained using any statistical package. It is important, however, to understand and remember the conditions.

$\overline{x}$ = sample mean

s = sample standard deviation

$Z_{\frac{\alpha}{2}}$ = the value from the Z-table with an area of $\frac{\alpha}{2}$ % to its right

$t_{\frac{\alpha}{2},n-1}$ = the value from the t-table with (n – 1) degrees of freedom (df) and an area of $\frac{\alpha}{2}$ % to its right

${\chi^2}_{\frac{\alpha}{2},n-1}$ = the value from the χ^2 table with (n – 1) degrees of freedom (df) and an area of $\frac{\alpha}{2}$ % to its right

df = Degrees of Freedom. It is defined as an unrestricted variable (i.e., free to vary) in a sampling distribution. This variable can take any positive integer value. Please refer Appendix A.3.4.4.6 “Degrees of Freedom” for more details.

Note: The second column provides the conditional requirements on which the selection of an appropriate CI formula should be based. The third column contains the CI formulas in inequality form, of which the left-hand expression represents the lower confidence limit and the right-hand expression represents the upper confidence limit.

Summary

CI for Population Mean:

- One-Sample Z: This is used when population standard deviation (σ) is known.
- One-Sample t: This is used when population standard deviation (σ) is unknown, sample size is small (< 30), and population is normally distributed.

CI for difference between two population means:

- Two-Sample t: This is used to determine the CI for the difference between two population means when the population standard deviations are the same but unknown.

CI for Population Standard Deviation:

- One-Variance: This is used to determine CI for the population standard deviation or population variance when the population is normally distributed.

2. Hypothesis Testing

A hypothesis is defined as an assumption about a population parameter. Examples of parameters are population mean and proportion.

Example of Hypothesis Testing:

- PQR Automobiles Inc., a leading car manufacturing company, is interested in knowing the effectiveness of its marketing campaigns on car sales.

  In order to test the effectiveness of marketing campaigns on car sales, the best possible approach is to collect the data before and after the campaigns. The data to be measured is the average (mean) number of car sales per week for a period of six months before and after the campaigns. The hypothesis can be, "There will be no significant change in car sales due to marketing campaigns." This hypothesis is a statistical comparison of means.

Hypothesis testing involves the careful construction of two statements: the null hypothesis and the alternative hypothesis.

- **Null Hypothesis**—This is the hypothesis about the population that needs to be verified. It is always about a population parameter and not about a sample statistic. The null hypothesis is denoted by H_o.

Example of Null Hypothesis:

- For the example given above, the null hypothesis is "There will be no significant change in car sales due to marketing campaigns." Statistically, it can be stated that there exists no statistically significant difference between the means.

- **Alternative Hypothesis**—This is the hypothesis to be accepted when the null hypothesis is not true. It is the opposite of the null hypothesis and is denoted by H_1 or H_a.

Example of Hypothesis Testing:

- Problem: To check if the mean is equal to 15

  Null hypothesis: H_o: Mean = 15

  Alternative hypotheses: H_1: Mean ≠ 15, or Mean > 15, or Mean < 15

Steps involved in testing a hypothesis

The steps involved in hypothesis testing are as follows:

1. Formulate the null and the alternative hypotheses.
2. Determine the α-value (significance level), which is similar to the use of α in confidence intervals. Normally a predetermined confidence level is given for a test.
3. Identify the test statistic applicable for the hypothesis test and calculate its value. Some of the inputs to the formulas come from the sample data.
4. Determine the critical values, typically found in tables such as Z, t, F, or χ^2. Use these values to define the critical region, also called the "reject region"—the set of values of the test statistic for which the null hypothesis is rejected.
5. Determine whether the null hypothesis should be rejected at a predetermined level of significance (α).
6. If the value of the test statistic falls in the reject region, then the null hypothesis is rejected and the alternative hypothesis is accepted.
7. If the value of the test statistic does not fall in the reject region, then the null hypothesis is not rejected.
8. State the inference in terms of the original problem.

Errors in Testing a Hypothesis

While validating any assumption about the population parameter or statistical distribution of the population using the hypothesis testing tool, two types of errors can occur:

- Type I Error: Also known as an error of the first kind, an α error, or producer's risk, it is the probability of rejecting a null hypothesis when it is actually true.
- Type II Error: Also known as an error of the second kind, a β error, or consumer's risk, it is the probability of failing to reject a null hypothesis when it is false.

A primary objective for a test would be to reduce both these errors as much as possible. The problem with this objective is that both of these errors cannot be reduced simultaneously. If a Type I error is reduced, then a Type II error will be increased, and vice versa. The best possible way to counter the problem is to fix one error and then try to reduce the other. In practice, the Type I error is pre-fixed at a specific level, and the Type II error is minimized; this is how the power of a test is determined.

The following are a few terms involved in testing a hypothesis:

Definition of Terms

- Statistical test: A decision function that takes its values in the set of hypotheses.
- Region of acceptance: The set of values of the test statistic for which we fail to reject the null hypothesis.
- Critical region: The set of values of the test statistic for which the null hypothesis is rejected.
- Power of a test ($1 - \beta$): The probability of not failing to reject a null hypothesis when it is false.
- Significance level of a test (α): This is another way of referring to a Type I error.
- *p*-value: In statistical hypothesis testing, the *p*-value is the probability of obtaining a result at least as extreme as the test statistic, given that the null hypothesis is true. Note that one rejects the null hypothesis if the *p*-value ≤ the significance level.

The hypothesis test enables us to make an inference about the true population value at a desired level of confidence.

Types of Tests

Various tests are employed based on the nature of the data. There are two kinds of data collected. They are continuous data and discrete data.

For continuous data, there are four major types of statistical tests:

- Population mean equal to a specified value
- Two population means are equal or not
- Population standard deviation equal to a specified value
- Two population standard deviations (or variances) are equal or not

For discrete data, there are two major types of statistical tests:

- Proportion equal to a specified value
- Two proportions are equal or not

Common Test Statistics

This section discusses the common test statistics used for the previously mentioned types of tests. When there is continuous data, a one-sample *z*-test or *t*-test is used for testing if the population mean is equal to a specified value. The population should be normally distributed for these tests to be valid. A *z*-test is used when the population standard deviation is known; a *t*-test is used when the population standard deviation is unknown. The normality assumption is not required for a *z*-test if the sample size is large enough. See Appendix A.3.5, Sampling Distributions, for details on sample sizes.

For comparing two population means, a two-sample *z*-test or two-sample pooled *t*-test is used. The common assumptions are as follows:

- Observations obtained from both of the populations should be independent.
- Both of the populations should be normally distributed.

A two-sample *z*-test is used when both population standard deviations are known, while a two-sample pooled *t*-test is used when the population standard deviations are unknown, but they are equal to each other. If the sample sizes are large enough, the normality assumption is not required for a two-sample *z*-test.

If the observations are not independent, they are called paired data and are treated as a single entity. For example, a manager wants to know whether there is any significant improvement in customer satisfaction scores as a result of employee training. To conduct the experiment, observations are taken in the form of test scores of each employee before and after the training, and a test score of each employee's customer satisfaction scores is taken before and after the training. The observations here are not independent, so a paired *t*-test is appropriate, using the differences of the paired data as a single entity. The assumptions are the same as in a one-sample *t*-test.

In some cases, a researcher will need to test if a population standard deviation is equal to a specified value. This can be tested using a chi-squared test. Again, the population needs to be normally distributed for the test to be valid. When comparing two population standard deviations, the *F*-test is used under the following assumptions:

- Observations obtained from both of the populations should be independent.
- Both of the populations should be normally distributed.

Two specialized tests—ANOVA and Chi-squared test are described in detail below.

Analysis of Variance (ANOVA): Test of difference between two or more means of independent samples

The ANOVA technique is used when testing differences between two or more means of independent samples. There are two parts in ANOVA: one-way classification and two-way classification. This section will only consider one-way ANOVA. The one-way classification treats data with only one classification criterion.

One-way ANOVA Model

Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$

Alternate Hypothesis(H_1): $All\ the\ \mu_i\ are\ not\ equal, i\ = 1,2,3, \ldots k$

$k =$ Total number of independent samples

$\mu_i =$ Population mean corresponding to the i^{th} sample

The steps to be performed are as follows:

1. Calculate variance between the samples (mean square between samples).
2. Calculate variance within the samples (residual mean square).
3. Calculate *F* ratio using the formula $F = \frac{\text{mean square between samples}}{\text{residual mean square}}$
4. Compare the calculated value of *F* with the critical value of *F* at 5% level of significance corresponding to the $F_{k\,-\,1, n\,-\,k}$ distribution.
5. When the calculated value > critical value, the null hypothesis is rejected, and the alternative hypothesis is accepted, meaning that there are significant differences in means.

One-way ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares (SS)	Mean Square (MS)
Between Samples	$k - 1$	SS_B	$\frac{SS_B}{k - 1}$
Residual	$n - k$	SS_R	$\frac{SS_R}{n - k}$
Total	$n - 1$	SS_T	

$$F = \frac{SS_B/_{k\,-\,1}}{SS_R/_{n\,-\,k}}$$

The F Statistic has a F distribution with $(k - 1, n - k)$ degrees of freedom under the null hypothesis(H_0).

Example of One-way ANOVA (Single Factor ANOVA):

- A manufacturing company has three plants: X, Y, and Z. A random sample of scrap generated in pounds was collected for six days, the results of which have been collected in the following table. Do the plants differ in waste production?

	Waste (in lbs)		
Observation No.	Plant X	Plant Y	Plant Z
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67

Assuming equal variances across plants, test whether the waste produced are the same at a 5% level of significance.

The null hypothesis H_o is, "All means are equal." The alternate hypothesis H_1 is, "At least two of the means are different."

Step 1: Calculate the sum, number, and averages of different levels.

Level	N	Sum	Average
Plant X	6	474	79
Plant Y	6	444	74
Plant Z	6	396	66

For example, the sum of all six observations for plant X is (85 + 75 + 82 + 76 + 71 + 85) = 474

The average is obtained by dividing the sum by the number of observations. The average = $\frac{474}{6}$ = 79

Step 2: Calculate total number of observations (N) = 6 + 6 + 6 = 18.

Step 3: Calculate grand total (T) = 474 + 444 + 396 = 1314.

Step 4: Calculate correction factor (CF) = $\frac{T^2}{N} = \frac{1314^2}{18} = 95922$ *(Calculation Tips: use Excel function =1314^2/18).*

Step 5: Calculate total sum of squares (TSS) = Sum of the squares of all individual observations – CF.

= $(85^2 + 75^2 + 82^2 + \ldots. + 75^2 + 67^2) - 95922 = 946$

Calculation Tip: enter all 18 observations in a single column. Use the Excel function =number^2 to calculate the square of the first observation, 85, in the second column, and then copy and paste this formula to the other cells in the second column to get the squared values of the corresponding observations. Finally, take the sum of all these squared values and subtract 95922 from this to get the final result of 946.

Step 6: Calculate the sum of squares of factors (SS_{Factor}) = $\frac{474^2+444^2+396^2}{6}$ – CF = $\frac{474^2+444^2+396^2}{6}$ – 95922 = 516 (Use Excel functions mentioned earlier.)

Step 7: Calculate the sum of squares of error (SS_{Error}) = TSS − SS_{Factor} = 946 − 516 = 430. (SS_{Factor} and SS_{Error} mentioned in this example are the same as SS_B and SS_R mentioned in the one-way ANOVA table)

Step 8: Calculate the total df (degrees of freedom) = Total number of observations (N) – 1 = 18 – 1 = 17.

df of Factors = Total number of factors − 1 = 3 − 1 = 2

df of Errors = Total df − df of factors = 17 − 2 = 15

4

Step 9: Construct ANOVA table.

Draw a table having five columns: sources, SS (sum of squares), df (degrees of freedom), MS (mean squares), and F as shown below:

Sources	SS	df	MS	F
Factor	516	2	258	9
Error	430	15	28.67	

- Enter factor and error as sources in the first column.
- Enter the corresponding SS value obtained from steps 6 and 7 in the second column.
- Enter the corresponding df value obtained from step 8 in the third column.
- Obtain the fourth column by dividing the values in the second column by the corresponding values in the third column.
- Obtain the fifth column by dividing the MS of factor (258) by the MS of error (28.67).

Formulas:

MS = $\frac{SS}{df}$

F = $\frac{MS\ of\ Factor}{MS\ of\ Error}$

Calculated value of F statistic = 9

Significance level (α) = 0.05

Critical value of F = Table value of F distribution $F_{0.95}$ with numerator df 2 and denominator df 15 = 3.68

Interpretation: Because the calculated value of F > critical value of F, the null hypothesis (i.e., the average test scores are the same for all the three plants) is rejected. At 5% significance, the plants do not produce the same amount of waste.

Chi-squared (χ^2) Test

A chi-squared (χ^2) test is applicable when data are categorical or when data are obtained by count (i.e., data consist of frequencies). The applications of the chi-squared test in marketing research are given below:

- Test of goodness of fit – This is a test to verify whether it is reasonable to regard a random sample as coming from a particular specified distribution (i.e., whether or not a particular model provides a good fit to the data). In other words, the marketing researcher is going to test if there is any significant difference between an observed frequency distribution and a hypothetical frequency distribution. The researcher is interested in determining whether the observed sample corresponds with a Normal, Poisson, or Binomial distribution, and how well the observed pattern fits the expected pattern.

- Test of independence of two factors using a contingency table – A contingency table is a two-way table of counts obtained when sample items are classified using two categorical variables. Each variable may have two or more categories. It is a test to verify whether or not the two classification criteria are independent.

Goodness of Fit Test

Null Hypothesis (H_0): $Data\ has\ the\ specified\ distribution$

Alternate Hypothesis(H_1): $Data\ does\ not\ have\ the\ specified\ distribution$

Test statistic $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

O_i = observed frequency, E_i = expected frequency; i = 1,2,…n

The test statistic has a $\boldsymbol{\chi^2}$ distribution with n-1 degrees of freedom under H_0

To determine degrees of freedom, the standard rule (number of observations − 1) is used. However, one must subtract an additional degree of freedom for each population parameter that has to be estimated from the given sample. Also, if two or more cells are grouped because of small, expected frequency, the respective number of degrees of freedom must be subtracted. See Limitations of the chi-squared test, later in this section, for more details.

4

Example of Chi-squared Test:

- To test whether a six-sided die is fair, a suitable model is $P(X = i) = \frac{1}{6}, i = 1,2,3,4,5,6$, where X is the random variable denoting the number marked on the upward side of the cube. The die is rolled 240 times successively, which produces the following results:

X	1	2	3	4	5	6
O_i	44	33	47	32	36	48

Null Hypothesis (H_0): $number\ thrown\ has\ the\ theoretical\ distribution\ P(X = i) = \frac{1}{6}, i = 1,2,3,4,5,6$

Alternate Hypothesis(H_1): $number\ thrown\ does\ not\ have\ the\ distribution$

Under the null hypothesis, each face of the die will be expected to appear $240 \times \frac{1}{6} = 40$ times; that is, the expected frequency is 40 for all the 6 cells.

Calculation of chi-squared statistic:

X	Observed Frequency (O)	Expected Frequency (E)	$O - E$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
1	44	40	4	16	0.4
2	33	40	-7	49	1.225
3	47	40	7	49	1.225
4	32	40	-8	64	1.6
5	36	40	-4	16	0.4
6	48	40	8	64	1.6
Total					6.45

Number of degrees of freedom = 6 − 1 = 5

Critical value = the upper 5% point of the χ_5^5 distribution = 11.07

Since the calculated value of the test statistic is less than the critical value, there is insufficient evidence to reject the null hypothesis at 5% level. There is not enough evidence to suggest that the die is not fair. Therefore, the null hypothesis can stand—the die is fair.

Contingency Table

To construct a contingency table, use two categorical variables X and Y. Assume that X has *r* number of categories and Y has *c* number of categories. Thus, the table will be a matrix with *r* rows and *c* columns denoted as an $r \times c$ contingency table.

$H_0: the\ two\ classification\ criteria\ are\ independent.$

$H_1: the\ two\ classification\ criteria\ are\ not\ independent.$

$Test\ statistic(\chi^2) = \sum \frac{(O_i - E_i)^2}{E_i}$, the summation is taken over all cells in the $r \times c$ contingency table.

The expected frequency for any cell is calculated using the formula : $E_i = \frac{row\ total \times column\ total}{table\ total}$.

The test statistic has a χ^2 distribution with (r - 1) (c – 1) degrees of freedom under H_0.

Example of Contingency Table:

- In a survey testing the effectiveness of an ad campaign encouraging the wearing of helmets as a safety measure while riding a motorcycle, 316 accident victims were classified according to the severity of their injuries, and whether they were wearing helmets at the time of the accidents. (Not wearing a helmet is equivalent to not seeing the ad.) The researcher must determine if the severity of injuries sustained is dependent on whether or not the victim has seen the ad.
  It is very clear that there are two factors to consider: severity of injury and presence of helmet. The survey needs to determine whether the severity of injuries sustained is dependent on whether the victims were wearing helmets. Note that the factors here are categorical variables with 3 and 2 categories, respectively.

  $H_0: severity\ of\ injuries\ is\ independent\ of\ wearing\ helmets.$

  $H_1: severity\ of\ injuries\ is\ not\ independent\ of\ wearing\ helmets.$

  3 × 2 Contingency Table:

	Wearing a Helmet	Not Wearing a Helmet	Row Sum
Death	5	53	58
Severe Injury	83	37	120
Minor Injury	108	30	138
Column Sum	196	120	316

Calculation of expected frequencies:

	Wearing a Helmet	Not Wearing a Helmet	Row Sum
Death	$\frac{58 \times 196}{316} = 35.97$	$\frac{58 \times 120}{316} = 22.03$	58
Severe Injury	$\frac{120 \times 196}{316} = 74.43$	$\frac{120 \times 120}{316} = 45.57$	120
Minor Injury	$\frac{138 \times 196}{316} = 85.59$	$\frac{138 \times 120}{316} = 52.41$	138
Column Sum	196	120	316

Calculate value of the chi-squared statistic:

$$\chi^2 = \frac{(5 - 35.97)^2}{35.97} + \frac{(53 - 22.03)^2}{22.03} + \frac{(83 - 74.43)^2}{74.43} + \frac{(37 - 45.57)^2}{45.57} + \frac{(108 - 85.59)^2}{85.59} + \frac{(30 - 52.41)^2}{52.41} = 88.27$$

The number of degrees of freedom = (3 – 1)(2 – 1) = 2

Critical value = upper 0.5% point of the x^2 distribution = 10.60

Because the calculated value of the test statistic is far more than the critical value, there is sufficient evidence to reject the null hypothesis at 0.5 percent level. Thus, it is reasonable to conclude that the severity of injury is almost certainly dependent on whether the victim has seen the ad.

Limitations of the chi-squared test

- A chi-squared test is sensitive to small expected frequencies in one or more number of cells in the table. The formula of the chi-squared test statistic consists of expected frequencies in the denominator of the expression, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$. Thus, if the expected frequencies in one or more cells in the table are very small, then the calculated value of the test statistic will be very high, potentially leading to an incorrect conclusion. It is therefore a rule that if expected frequencies in one or more cells in the table are less than 5, then those cells are grouped together to get a revised expected frequency greater than or equal to 5. The loss in degrees of freedom will be equal to the number of cells reduced. If 3 cells have an expected frequency less than 5, those cells are grouped into one cell. So, the total number of cells reduced = 3 – 1 = 2. Two degrees of freedom are subtracted.

- A chi-squared test does not provide much information on the degree of association or the direction of a relationship.
- A chi-squared test is sensitive to sample size. The size of the calculated chi-squared is directly proportional to the sample size. This may make a weak relationship statistically significant if the sample size is large enough.

4.2.2.2 Bivariate Data Analysis

In univariate analysis, researchers deal with only one variable or factor, allowing them to analyze the central tendency and dispersion of the data. In practical situations of marketing research, however, the researcher often needs to analyze data with two or more variables or factors. In these cases, the focus will shift to other statistical measures such as covariance and correlation coefficients.

Bivariate analysis involves analysis of two variables or factors simultaneously. Because bivariate analysis involves two variables, the researcher is interested in identifying the relationship between them (i.e., whether or not one variable is dependent on the other and the nature of the relationship). Instead of measuring central tendency and dispersion or variance, as in the case of univariate analysis, the researcher will be measuring the degree of linear relationship between the variables (i.e., the correlation coefficient).

1. Correlation and Regression Analysis

When dealing with two or more variables, any existing relationships between or among them and the impact of one variable on the other(s) can be studied using correlation and regression analysis.

Correlation

Correlation examines the relationship between two variables using a correlation coefficient measure that indicates the strength of a relationship.

For example, consider two variables, X and Y. The objective is to find out whether or not any relation exists between them. If any relation exists between them, for example $Y = f(x)$, then the researcher must define the relationship f in the best possible way.

X is called an explanatory or independent variable, and Y is called a response or dependent variable.

Correlation Coefficient

When determining the priority or urgency of a relationship between a defect and an identified cause for that defect, the researcher uses a correlation coefficient.

A correlation coefficient is a measure of strength of a linear relationship between two variables denoted by r with the given range $-1 \leq r \leq 1$.

If the value of r is 1 or close to 1, then the two variables have a strong positive relationship.

If the value of r is -1 or close to -1, then they have a strong negative relationship.

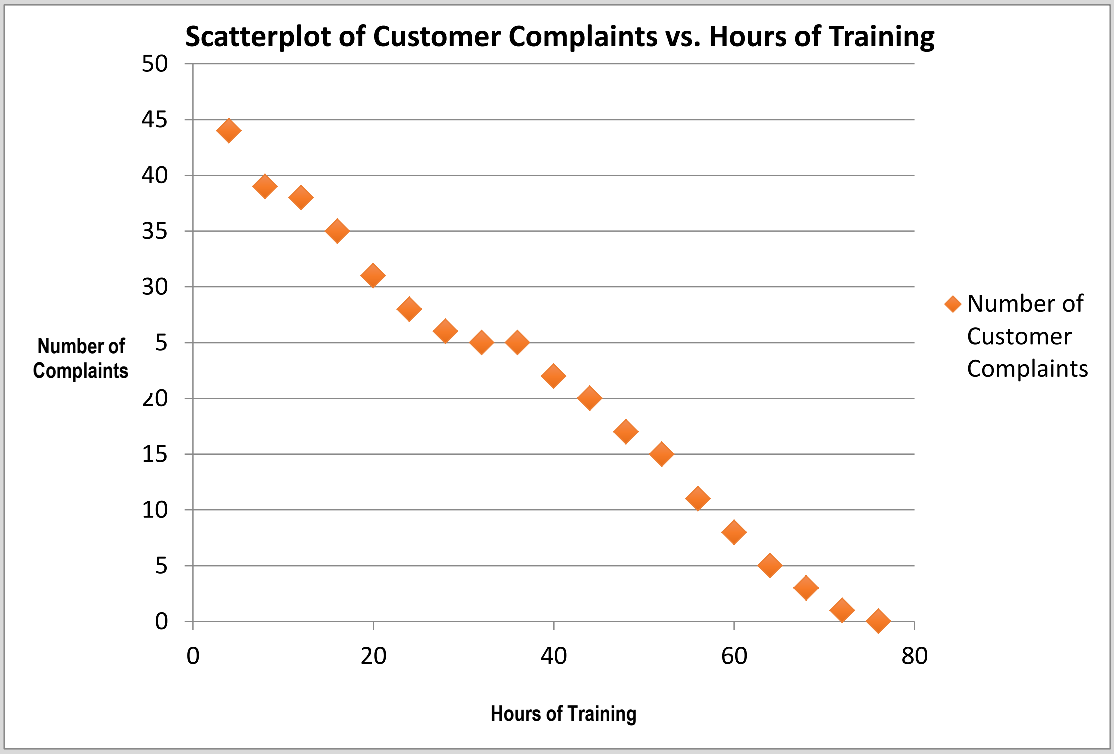
If the value of r is 0, then no relationship exists.

Example of Correlation Coefficient:

- One of the team leaders at ABC Inc. decides to define the impact of training on customer service performance. The following data are collected:

Hours of Training	Number of Customer Complaints
4	44
8	39
12	38
16	35
20	31
24	28
28	26
32	25
36	25
40	22
44	20
48	17
52	15
56	11
60	8
64	5
68	3
72	1
76	0

To fulfill the requirement, the team leader needs to draw the scatter plot and identify the relationship between variables (training hours and number of complaints).



Interpretation

From the scatter plot, it is clear that the hours of training and number of customer complaints logged have a strong negative relationship. This means that as the number of training hours increases, the performance level also increases. In other words, the more training the staff members receive, the fewer complaints, which leads to increase in performance.

The correlation coefficient r = –0.995. Note that the value of “r” is very close to –1.

A few important points to remember:

- Correlation analysis is used to determine the degree of linear relationship between two or more variables but does not infer any cause-and-effect relationship.
- Even a high degree of correlation does not necessarily imply existence of a cause-and-effect relationship between variables.
- Though correlation does not imply causation, the existence of causation always implies correlation.

The significant degree of correlation between two variables may be attributed to one or more of the following reasons:

- pure chance, particularly in small samples
- an influence of one or more other variables on both of the correlated variables
- both variables influencing each other so that neither can be considered the cause of the correlation

There are various methods of studying correlation including the scatter diagram method, Karl Pearson's coefficient of correlation, and the Rank correlation method. These, and several others, are examined below.

- **Scatter Diagram**

  A scatter diagram is a graphical representation of the relationship between two variables, and it provides both a visual and statistical means to test the strength of a relationship. A scatter diagram is also referred to as a scatter plot.

  To construct a scatter diagram, the data is collected for both variables (preferable sample size of 20 or more) and the data points are plotted on an XY plane such that variable 1 is plotted along the X axis and variable 2 is plotted along the Y axis.

- **Karl Pearson's Coefficient of Correlation (r)**

  This method is the most popular way to study correlation between two variables. This method is applicable when data is measured in interval or ratio scales.

- **Rank Correlation Method**

  This method was developed by the British psychologist Charles Edward Spearman. This method is used for categorical data or ordinal data. There may be some factors for which quantitative measures are not possible (data is not continuous) but the data can be arranged in some order such that each individual is allocated a rank. The Spearman's rank correlation coefficient is calculated based on rank order.

  The choice of correlation coefficient depends on the variables for which the correlation coefficient needs to be calculated. The Karl Pearson method for calculating a correlation coefficient is based on certain assumptions (the contributed data have to be measured in interval or ratio scale and the variables should have a linear relationship that can be checked by plotting a scatter diagram, as discussed earlier). On the other hand, correlation coefficient calculated using the Spearman method is a non-parametric or distribution-free process (i.e., it does not require the distributional and linearity assumption of the variables). The Spearman method is more robust than the Pearson method when the distributional assumptions are not met. Therefore, when the Karl Pearson method cannot be used (i.e., variables are measured using an ordinal scale and the underlying distribution of data is not known), the best possible alternative is Spearman's rank correlation method.

Regression Analysis

Regression analysis examines the relationship between one dependent variable and one or more independent variables. When there is only one independent variable it is said to be a simple linear regression.

- **Simple Linear Regression Model**

  A linear model is defined as a relation between two variables where changes in one variable produce a proportionate change in the other variable. Mathematically, a linear model is expressed as $Y = a + bX$, where a and b are constants that need to be estimated from the data. The strength of the linear relationship between two variables is detected with the use of a scatter plot and correlation coefficient. After that, the linear model to be used in forecasting is determined. See appendix A.3.6 to learn how the regression line is derived using the least square method.

- **Nonlinear Relationship between Two Variables**

  When two variables do not have a linear relationship (which is the case in many practical situations), their relationship can be converted into a linear model with suitable transformation.

  For example, let two variables X and Y have a relation such that $Y = ab^X$.

  Taking the logarithm on both sides produces $\log Y = \log a + X \log b$. This can be written as $Y_1 = a_1 + b_1 X$, which is a linear model in X and Y_1.

  Further analysis can be done using this linear model.

Various Types of Regression

- Ordinary Linear Regression—This is the most basic type of regression between a predictor(s) and response variable. It models the linear relationship between predictor(s) and response when all data is metric and the response variable is continuous.
- Nonlinear Regression—This models the relationship between predictor(s) and response when quadratic or cubic terms are not adequate. It is used for nonlinear relationships, such as nonlinear growth or decay.
- Orthogonal Regression—This models the relationship between one predictor and one response when the measurements of both the response and predictor include random error.
- Binary Logistic Regression—This models the relationship between predictor(s) and a response that is a binary or dichotomous variable.
- Ordinal Logistic Regression—This models the relationship between predictor(s) and a response that has three or more outcomes with an order (i.e., high, medium, and low).
- Nominal Logistic Regression—This models the relationship between predictor(s) and a response that has three or more outcomes that do not have an order (i.e., red, blue, green).

A more practical situation occurs when the researcher has two or more independent variables or predictors. This case will be discussed next.

2. Two-way ANOVA

The previous section discussed one-way ANOVA. ANOVA can be applied in a two-way classification. In this classification, data are classified according to two criteria of factors. In addition, when the researcher is conducting two-way ANOVA, it can be done with or without replication. In the case of two-way ANOVA without replication, the researcher is measuring two or more factors and running only one instance of the sampling. In the case of two-way ANOVA with replication, the researcher is running two or more factors and has two or more samples for each data point.

Example: Two-way ANOVA with replication:

- A hotel is using two factors (price and city) to determine their significance on the customer satisfaction score. For each price point (Low, Medium and High) the researcher samples 6 times in each city (therefore, with replication). The data in the table are the customer satisfaction (CSAT) scores on a 100 point scale.

City	Low (20)	Med (30)	High (40)
London	77	74	49
London	71	76	54
London	68	69	56
London	90	66	59
London	72	66	55
London	81	74	67
Madrid	90	89	85
Madrid	78	64	60
Madrid	75	70	65
Madrid	89	79	65
Madrid	94	89	59
Madrid	100	96	83
Moscow	88	73	61
Moscow	71	43	40
Moscow	72	48	42
Moscow	80	49	39
Moscow	94	84	65
Moscow	86	60	50
San Francisco	100	92	81
San Francisco	76	56	45
San Francisco	87	51	39
San Francisco	76	57	43
San Francisco	80	66	53
San Francisco	109	100	87

The data are classified according to two criteria of factors: city and price. The researcher is interested in understanding if these factors had a significant effect on the satisfaction scores and whether the interaction of these two factors (city X price) had a significant effect. Refer appendix A.3.7.2 for a detailed analysis.

4.2.2.3 Multivariate Data Analysis

A more practical and complex method is needed when the researcher has more than two variables to analyze. The analysis involving more than two variables or factors is called Multivariate Analysis. The most favorable scenario in this case is to have one dependent variable and multiple independent variables (i.e., one particular factor that is influenced by multiple other factors). Another scenario could have two or more variables that are dependent on multiple other independent variables. These two scenarios require multivariate dependence techniques. Based on the number of dependent variables and type of data (categorical or metric), there are several multivariate dependence techniques, including the following:

- Multiple Linear Regression
- Discriminant Analysis

In multivariate interdependence techniques, there exist no dependent or independent variables. Here, all of the variables are treated equally to discover the interdependent relationship. Among the multivariate interdependence techniques are the following:

- Factor Analysis
- Cluster Analysis
- Conjoint Analysis

1. Multivariate Dependence Techniques

- **Multiple Linear Regression**

In cases when multiple linear regression is appropriate, there will be two or more explanatory or independent variables and one dependent variable. The dependent variable must be a metric (data measured on an interval or ratio measurement scale). These are the most common practical situations market researchers come across.

Example of a dependent variable:

- For example, the price of a home is attributed to more than one factor: dimensions of the plot, number of bedrooms, number of floors, type of building, location of property, and so on. Here, the price of a house is the dependent variable: it is dependent on a number of other independent variables.

The general form of the multiple linear regression model is as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Where x_1, x_2, …, x_k are k independent variables, y is the dependent variable, and $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients.

The coefficients can be determined by the same method (method of least square) used in the simple linear regression model i.e. by minimizing the residual sum of squares (SSR).

Basic assumptions for the Multiple Linear Regression model are as follows:

- Linearity—A linear relationship should exist among the independent variables (also known as predictors) and between the predictors and the dependent variable.
- Independence—The error terms or residuals are independent (no serial correlation). The predictor variables must be statistically independent and uncorrelated to one another.
- Normality—The error terms are independent and identically distributed as N(0, σ^2) variable.
- Homoscedasticity—The variance of the error variables is constant.

Test for Significance of the Overall Model

Null Hypothesis (H_0): None of the predictor variables are important for predicting *Y*.

H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$.

Alternate Hypothesis (H_1): At least one of the predictor variables is important to predict *Y*.

H_1: At least one $\beta_i \neq 0$.

Test Statistic: $F = \frac{Regression\ MS}{Residual\ MS} = \frac{Regression\ SS/k}{Residual\ SS/(n-k-1)}$, which follows *F* distribution with (k, $n-k-1$) degrees of freedom, n = total number of observations, and k = number of independent variables.

Regression SS = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$.

Residual SS = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

These calculations can be easily obtained from the ANOVA table as an output for any statistical tool.

The ANOVA table displays the calculated value of the *F*-statistic along with the *p*-value. A small *p*-value (conventionally less than 0.05) rejects the null hypothesis i.e. if the *p*-value is very small (<0.05) the researcher can conclude that at least one of the independent variables is significant.

Once this conclusion is reached, the researcher can test each predictor variable separately.

Test for Individual Terms

Once it has been concluded that at least one of the predictor variables is significant in predicting the response *Y*, the researcher needs to check how many of the *k* predictors are statistically significant so that the non-significant predictors can be removed from the model.

If a predictor variable does not contribute significantly to predict Y, the coefficient of that variable should be zero. Therefore, the researcher is interested in testing the hypothesis that the estimated coefficient corresponding to a predictor variable is significantly different from zero.

Framing of Hypothesis:

H_0: $\beta_i = 0$ (i.e., the independent variable x_i is not significant in predicting Y)

H_1: $\beta_i \neq 0$

$i = 1, 2, \ldots, k$

Test statistic: $t = \frac{\widehat{\beta_i}}{se(\widehat{\beta_i})}$ which follows a t distribution with $(n - k - 1)$ degrees of freedom.

Please note that this test is the same as the test for significance in simple linear regression.

The calculations can be easily obtained from the parameter estimates table that is an output for any statistical package. The table will display the estimated value of the regression coefficients, standard error of the estimates, value of the t-statistic, and p-value for each predictor variable. A small p-value (less than 0.05) will reject the null hypothesis; that is, if the p-value is less than 0.05, then the corresponding predictor variable is statistically significant.

The non-significant predictors can be discarded just by looking at the corresponding p-values in the parameter estimates table.

Strength of Association

Strength of association is a measure using the coefficient of determination (r^2) similar to the case of simple linear regression. This value can be obtained easily as an output for regression analysis for many statistical tools.

Another measure called r^2 adjusted will be available along with the r^2 value. This is used to adjust the r^2 statistic for addition of non-significant predictors to the regression model. Thus, for simple linear regression, r^2 adjusted is almost equal to r^2. The formula is as follows:

$$r^2\ adjusted = r^2 - \frac{k(1 - r^2)}{n - k - 1}$$

The value of r^2 should be close to 1 for the model to be appropriate. Also r^2 adjusted should be close to r^2.

Multicolinearity

Multicolinearity is defined as the presence of a high degree of correlation between several independent variables in a multiple regression model. It can produce unexpected results; for example, the model may fit the data well (high value of F statistic) even though none of the predictor variables are statistically significant in predicting Y.

Multicolinearity increases the standard error of the coefficients, which in turn leads to a small observed t statistic. This can allow too many null hypotheses to be accepted. In other words, multicolinearity makes some predictor variables insignificant when they should be otherwise significant.

Variance Inflation Factor (VIF) is a measure of multicolinearity in a multiple linear regression model. Most advanced statistical tools produce VIF as an output of the regression analysis. VIF quantifies how much variance of the estimated coefficient is inflated due to the presence of multicolinearity.

Conventionally, if the VIF is > 10 for a predictor variable, then colinearity associated with that variable exists. If two or more predictor variables have VIF > 10, one of them must be removed from the model. The best way to select which variable to remove is to remove each of them one at a time, test the model without that variable, and then select the model with the highest r^2 value.

Checklist for Multiple Linear Regression:

- Check linearity. (A scatter plot is useful to check linearity.)
- Test for significance of the overall model (*F* test).
- Check whether the value of r^2 is reasonably high.
- Test for significance of each predictor (*t* test) and remove any insignificant predictor from the model.
- Check if the value of r^2 adjusted is close to r^2.
- Check normality. (A histogram of the residuals is useful to check normality.)
- Check homoscedasticity. (A scatter plot of the residuals with each predictor variable or a scatter plot of residuals against the predicted values is useful to check homoscedasticity.)
- Check multicolinearity. (Check for the VIF value for each predictor variable. If the VIF > 10 for some predictor variables, then it indicates the presence of multicolinearity. Remove those predictors that have a VIF > 10 one at a time and select the model with the highest r^2 value.)

Example of Multiple Linear Regression:

- The following table provides data points obtained from fifteen randomly selected trainees/ delegates of a professional certification and training provider. The organization is a leading global certification provider offering certification courses in all formats i.e. online with proctored exam as well as traditional classroom mode. The data is a random sample collected from the feedback given by the delegates taking certification courses.

Customer satisfaction score	Hours spent in class training	Hours spent in online training	Number of practice tests	Number of customer referrals
85	56	22	3	2
69	43	20	1	2
55	39	15	1	1
92	65	34	4	5
70	44	25	2	3
68	40	20	2	2
81	45	27	2	3
60	32	18	1	1
78	39	24	2	2
55	31	18	0	1
74	46	24	1	2
65	38	14	1	2
78	50	21	0	2
94	62	35	3	3
66	48	17	1	3

It is clear from the data given that the factors which affect the output "Customer satisfaction scores" are, hours spent in class training, hours spent in online training, number of practice tests, and number of customer referrals. Because there is more than one predictor variable, the researcher will need to use a multiple linear regression model. See Appendix A.3.7.3 for detailed analysis using Minitab.

Drawbacks of Multiple Linear Regression

This type of regression is the most basic type and easily calculable, but it is applicable only with metric data (i.e., where the response variable or dependent variable is continuous). Thus, it is not suitable for predictive analytics using continuous data. Therefore, a modern regression technique is used when the response variable is binary. This is known as logistic regression. A more robust form of logistic regression is logic regression. These techniques are useful in customer retention, for example, detecting which variables are more or less predictive of creating a desired customer outcome.

- **Discriminant Analysis**

Discriminant analysis is similar to multivariate regression analysis in terms of usage. Both of these techniques are useful in forecasting or predictive modeling. The only difference between them is that the dependent variable is continuous in regression analysis, but for discriminant analysis the dependent variable is a categorical variable. Market researchers use discriminant analysis when there is only one dependent variable that is categorical and two or more independent variables. The dependent variable can have two or more categories. If the dependent variable has only two categories, it is called dichotomous.

Example of Discriminant Analysis:

- ABC Inc., a leading Internet broadband service provider headquartered in Washington, DC, is interested in knowing the portion of its customer base that intends to leave the network based on factors such as demographics, customer annual income, and customer feedback. By segmenting the existing customer base, the company can design its marketing strategy for customer retention. In this case, the dependent variable, also called a grouping variable or criterion variable, is a dichotomous variable as it classifies the existing customer base into two segments: one group that intends to leave the network and the other that will continue subscribing to the service.

In regression analysis, the objective of the study is to predict a value of the continuous dependent variable based on many continuous or discrete independent variables. In contrast, discriminant analysis is used in the classification or categorization of two or more groups based on a number of independent variables (predictor variables). The independent variables can be continuous or discrete.

Like the multiple linear regression model, discriminant analysis is a complex statistical technique that needs advanced statistical packages.

Discriminant analysis has many applications in marketing. It can be used in the banking and financial services industry to predict potential defaulters for products such as loans and credit cards before such an incident actually happens. Discriminant analysis can be used by a company to classify its customer base into potential buyers and non-buyers for a new product.

2. Multivariate Interdependence Techniques

- **Factor Analysis**

Factor analysis is a technique where all the variables are treated equally (i.e., the concept of dependent or independent variables does not exist). It is an interdependence multivariate technique.

Unlike regression analysis or discriminant analysis, factor analysis treats all the variables equally. The main objective of factor analysis is to reduce a large number of variables into a relatively small number of factors by considering interdependence among all the variables simultaneously.

Marketing researchers use factor analysis as a data reduction tool when many observed correlated variables can be reduced to a potentially lesser number of unobserved and uncorrelated variables called factors. The theory behind factor analysis is that two or more correlated variables can be grouped into one factor and that the factor will simultaneously represent the features described by the observed correlated variables. Thus, factors are a linear combination of observed variables having high correlations. More simply put, if two variables are highly correlated (i.e., one can be explained by the other as has been shown in simple linear regression), these two variables can be grouped to get a single variable—often called a hidden variable or unobserved variable—which will be representative of the data described by the observed correlated variables. The same logic applies when there are multiple observed correlated variables.

Example of Factor Analysis:

- For a given data set with many variables, a correlation matrix is produced to see how many variables have high correlations. A correlation matrix is a matrix whose elements are correlation coefficients between two variables. Here are five hypothetical variables for which a correlation matrix is drawn.

	V1	V2	V3	V4	V5
V1	1	0.78	0.06	-0.12	0.65
V2	0.78	1	-0.21	0.08	0.83
V3	0.06	-0.21	1	0.79	0.18
V4	-0.12	0.08	0.79	1	-0.1
V5	0.65	0.83	0.18	-0.1	1

Each cell of this matrix represents the correlation coefficient between the corresponding variables. It is easy to notice that the diagonal entries of the matrix will be 1 and entries above the diagonal are the same as the corresponding entries below the diagonal. This is because the correlation coefficient between variables 1 and 2 is the same as the correlation coefficient between variables 2 and 1. This kind of matrix is called a symmetric matrix. The variables having comparatively high correlations in the matrix can be combined. Variables 1, 2, and 5 can be grouped together to form a factor A because the variables 1, 2, and 5 have relatively high correlations among themselves. Similarly, variables 3 and 4 will form another factor B. Thus, five variables are now reduced to two uncorrelated factors.

The data reduction or summarization achieved by reducing a number of observed correlated variables to a relatively lesser number of uncorrelated factors helps marketing researchers to identify hidden factors in product characteristics and customer preferences. Factor analysis is also helpful in the sense that it summarizes a large data set into a relatively smaller data set—without losing any critical information—which makes further analysis easier.

- **Cluster Analysis**

Cluster analysis is another interdependence multivariate technique used in marketing research. It is similar to factor analysis in the sense that both treat all of the variables equally and simultaneously to establish an interdependence relationship. The main difference between factor and cluster analysis is that cluster analysis reduces the number of objects, whereas factor analysis reduces the number of variables.

Cluster analysis is used to create groups of objects having similar characteristics. A large number of objects are divided into some groups. These groups are known as clusters. These clusters have certain properties. Within a cluster, the elements or objects are homogeneous in nature, and are heterogeneous among objects belonging to different clusters. The clustering process (i.e., grouping objects into different clusters) depends on the similarity of characteristics among the objects. For continuous variables, the similarity is measured using a distance metric (i.e., distance between two objects). The clusters are formed in such a way that the distances between objects within a cluster are normally small, while the distances between objects belonging to different clusters are large. In other words, clusters are composed of objects that are closer to each other. The proximity, or closeness, is based on the similarity of features among objects.

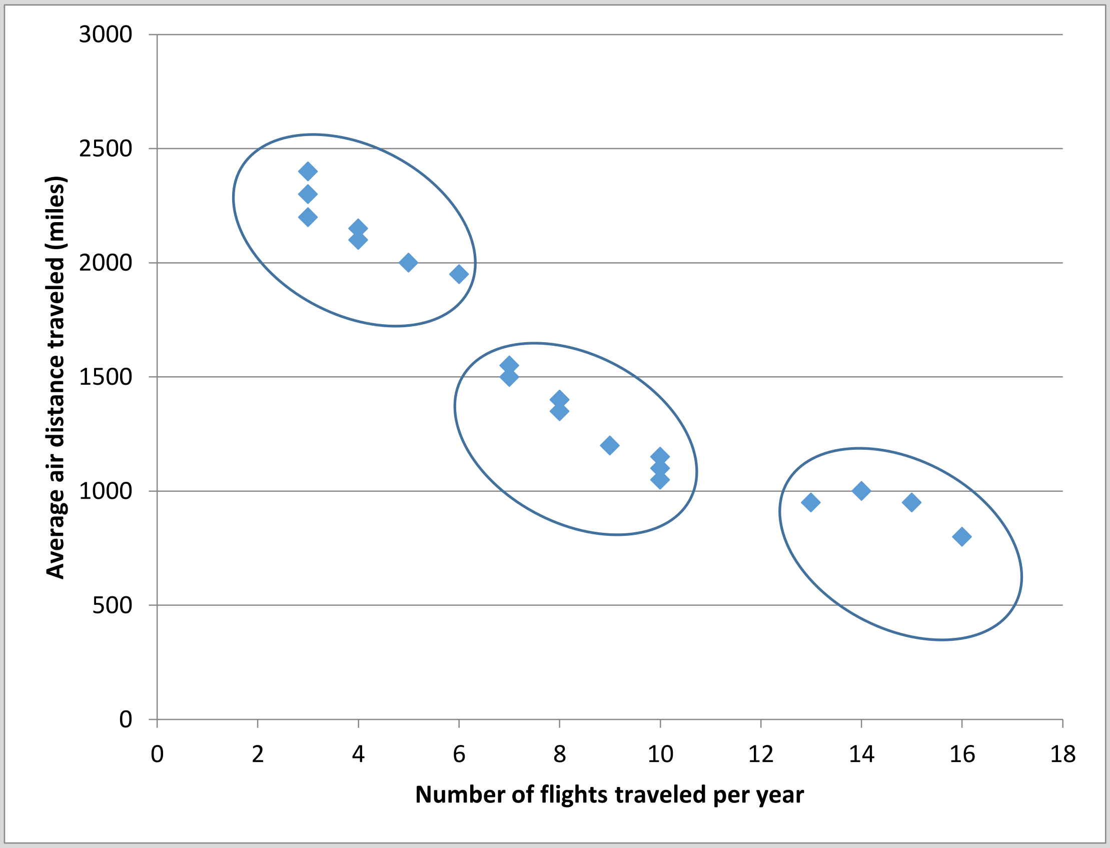
Market researchers use this technique to segment a customer base into different groups to implement effective marketing programs.

Example of Cluster Analysis:

- An airline company is analyzing its customer base using the data given below. The objective is to segment the customer base so that the company can develop targeted marketing programs for different groups within its customer base.

Customer Name	Number of Flights Traveled per Year	Average Air Distance Traveled (miles)
Armando	14	1000
Ben	10	1100
Claire	4	2150
Danielle	3	2200
Eugene	16	800
Frank	13	950
George	7	1500
Hannah	4	2100
Irene	8	1400
Jorge	15	950
Karen	8	1350
Liam	3	2400
Madeline	8	1400
Noah	9	1200
Oprah	10	1150
Peter	6	1950
Quentin	3	2300
Rosario	7	1550
Sonia	5	2000
Talluah	10	1050

A simple scatter diagram is drawn with these 20 objects. This is a simple case because there are only two variables or characteristics involved. In a practical scenario, there may be multiple variables involved, for which the market researchers consider using advanced statistical tools such as SPSS, SAS, or Minitab.



In this diagram, each point represents a customer (object). The horizontal axis represents the number of flights taken per year and the vertical axis represents average air distance traveled (miles). Here, the objects have two characteristics: the number of flights taken per year and the average air distance traveled (miles). The objects can be segmented into three groups based on the distance between objects. These groups are marked in the above diagram. All of the groups or clusters are formed in such a way that the objects within clusters are close to each other with larger distances between the clusters. In other words, the objects belonging to different clusters are separate from one another.

The cluster at the top contains objects C, D, H, L, P, Q, and S. The objects in this cluster are similar in nature because this cluster comprises all of the customers who travel a few times in a year, but travel long distances. The cluster in the middle contains objects B, G, I, K, M, N, O, R, and T. This cluster comprises all of the customers who travel a moderate number of times per year for moderate air distances. The remaining cluster contains objects A, E, F, and J. This cluster comprises all of the customers who travel many times a year for normally short air distances. These customers are frequent flyers who commute small distances. The first group can be thought of as travelers who are taking personal flights such as occasional vacations. The last group can be thought of as business travelers taking short business trips multiple times a year. The group in the middle could be a mix of business and personal travelers.

Once the customer base is segmented, a company can develop specific marketing programs for different target markets.

For example, in this case, the airline company could offer special discounted flights for its frequent flyers, such as monetary discounts, gift vouchers, free return air fare after completing a specified number of travels within a certain period, and free first-class lounge access. On the other hand, the company could offer special vacation discounts combined with accommodations at premium hotels for its infrequent travelers who normally travel long distances to international destinations.

Cluster analysis thus helps marketers concentrate on a specific target market and develop marketing programs tailored to the characteristics of that target group. This will enhance the effectiveness of the marketing programs implemented by the company.

- **Conjoint Analysis**

Conjoint analysis is an advanced multivariate technique used extensively in marketing research. Customers purchase a product or pay for a service after carefully inspecting the various features or the combination of features a product or service offers. Thus, the value a customer is willing to pay for a product or service is dependent on the proper blend of attributes at desired levels. Conjoint analysis is a multivariate technique that quantitatively analyzes the joint effect of multiple attributes or features of a product or service that a customer desires.

Example of Conjoint Analysis:

- John is interested in purchasing a bicycle for his school-aged child. He will probably not limit his search to one attribute; instead, he will be interested in a combination of attributes for a given price range. The attributes could be color, height, weight, wheels, and gearing. For each attribute, there are various levels that he will consider. The levels for the attribute color are the standard bicycle colors available in the market (e.g., red, blue, green, yellow, pink, black, and grey). The levels for height could be low, medium, or tall. Similarly, there will be levels for other attributes. There are many combinations of attributes and corresponding levels. With the help of conjoint analysis, it is possible to identify the best possible combination for which the customer is willing to pay a certain price.

When a company plans to launch a new product or service, the pricing strategy is determined based on the conjoint analysis. Market researchers use this technique to fix the required price for a product having a blend of features that a customer desires. Market researchers typically frame a survey questionnaire known as a conjoint survey.

The starting point for the survey is to consider all the possible combinations of attributes and levels. It is advisable not to consider too many attributes and levels because doing so will make the number of combinations too high to analyze. To finalize the required attributes for a product or service, the market researcher may interview customers, conduct a focus group, or interview the product managers who understand the product features completely. Once the attributes and their respective levels are finalized, the possible combinations are created.

Each combination is a set of attributes with associated levels. Each such combination is written on a card with details such as product description, pictures, and prices. Each customer is provided a pack of cards. Customers are then requested to provide a rank order for all of the cards on the basis of their preference. Suppose there are sixty combinations of attributes and levels for a product. Sixty cards, each representing one combination, are given to a customer who is instructed to arrange the cards in order of preference or willingness to purchase. Sometimes the ranking method becomes a little difficult for some respondents when there are many cards. Respondents may instead use a 1–10 Likert scale to rate each combination. (Note that when a ranking method or rating scale is adopted for collecting data from a selected class of respondents, it should be consistent for all respondents.)

The next step is to aggregate the collected responses for further analysis. It is very tempting to take the averages of the responses to get a consolidated ranking or rating. But this may lead to an incorrect result if the class of respondents is not homogeneous. Therefore, before taking averages of the ranking or ratings for each respondent, the market researcher must perform a cluster analysis to segment the target customer base or respondents.

Conjoint analysis is a very important tool for conducting primary market research to understand consumer preferences before launching a new product or service in the market. This is particularly important in large scale manufacturing industries where a manufacturer invests a lot of money and time in producing outputs (products). Launching a product in the market without taking consumer preferences into consideration can result in poor sales. As a consequence, the manufacturer has to suffer a huge financial loss. This situation could have been avoided if a proper conjoint analysis to understand consumer preferences was performed before launching the new product.

4.2.2.4 Statistical Packages*

As long as the amount of data is small, the researcher or statistician can perform data analysis manually. Though this can provide the desired result, doing statistical analysis manually is time consuming. Real-life marketing research problems are usually based on huge amounts of data for which manual processes of statistical analysis are practically impossible. Thus, researchers use advanced statistical packages for various types of statistical analysis of large data sets. These packages are specialized computer programs for statistical analysis. Some of the packages require a code or program to be written to execute a statistical analysis, whereas other advanced packages are made very user-friendly and do not require any coding. These packages have built-in programs, which run at the backend to generate the outcome of any statistical tool, such as regression, ANOVA, and the chi-squared test.

Most of these packages have similar features:

- A large data set can be analyzed.
- Most statistical tools/concepts/topics are covered.
- The outputs of most of the statistical analyses consist of graphs, charts, and plots.

These packages differ from one another in factors such as interface, storage/data capacity, user friendliness, and knowledge and coding required to execute a statistical analysis or use a menu tool bar/dialog box commands. These packages make statistical analysis of a large data set very easy and significantly faster than manual processes, as the researcher needs to know only the tool(s) to be used for a given scenario and how to interpret the results. Refer Appendix A.3.7.

Example of Statistical Packages:

- Some of the popular statistical packages used in marketing research are SPSS, SAS, R, and Minitab.

4.2.3 Outputs

4.2.3.1 Analyzed Data*

Once the raw data is processed and prepared for analyses, the data analysis is completed using the research design that has been established in the *Choose Research Design* process. It is important that the relevant and the right amount of data is processed in order to ensure the analysis process is efficient and yields useful information. Data analysis involves converting processed data into meaningful information that the marketing team can use to make key decisions. Deriving valuable information from the *Data Analysis* process requires the use of various key tools including estimating the values of unknown parameters and testing hypotheses using a variety of tools for drawing inferences. Data analysis is a key process in developing answers to research problems. The resulting analyzed data is an input for interpretation and research reporting.

5. DATA INTERPRETATION AND REPORTING

This chapter deals with interpreting and reporting on the analyzed data from the Marketing Research project. This is an important step in transforming the analyzed data into meaningful and reliable information that can be used to solve the research problem and ultimately inform key marketing and business decisions.

5.1 Data Interpretation—In this process, all of the possible interpretations from the analyzed data are determined. These interpretations help in reporting the results of the Marketing Research.

5.2 Reporting—In this process, guidelines are provided to create the Marketing Research report based on the data interpretation and the research problem and objectives. Subsequently, specific recommendations are established for consideration by senior management.

Figure 5.1 provides an overview of the processes described in this chapter.

5.1 Data Interpretation	5.2 Reporting
INPUTS 1. Analyzed Data* 2. Research Problem and Objectives* **TOOLS** 1. Tables* 2. Charts* 3. Expert Judgment **OUTPUTS** 1. Interpretations*	**INPUTS** 1. Interpretations* 2. Research Problem and Objectives* 3. Senior Management Direction and Insights **TOOLS** 1. Oral Reporting 2. Report Writing* 3. Report Format 4. Presentation Software **OUTPUTS** 1. Research Report* 2. Recommendations

Figure 5-1: Data Interpretation and Reporting Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

5.1 Data Interpretation

Interpreting the marketing research data includes understanding the output of the Data Analysis process. For easy reference, information is usually displayed as data tables. In data tables, results can be displayed in both numerical and percentage formats.

Marketing researchers identify changes in data patterns while interpreting the data. These changes are more apparent for survey results taken from multiple time periods. For example, customers might be asked to rate their experiences at a small restaurant on cleanliness, quality of food, timeliness, ambience, and responsiveness. Improvements may be needed in all these sections, but the researcher might observe a significant decline in cleanliness ratings over a period of time. Therefore, the interpretation can be that cleanliness is the main area of concern for the restaurant and needs improvement.

Also, proper understanding of research objectives is very important in interpreting results, as the research objectives should guide the focus of the interpretation effort.

Example of Data Interpretation:

- A survey was conducted to understand the role of influencers on consumers shopping for an automobile insurance policy. The hypothesis was that the most significant influence came from peer groups when choosing an insurance company. The data interpretation results proved that the peer group influence differed significantly for 18–35-year-olds vs. 36–55-year-olds.

  In order to improve the market share in the category, insurance companies would need to explore, for 36–55-year-olds, programs that supported referrals from peer groups (both personal and professional), while with younger categories it was observed that the single largest influence was their parents, followed by friends. In both cases, building referral programs that leveraged the influence groups was the goal.
  By reviewing the survey data, user patterns of behavior became apparent. For example, the data showed that there are consistent patterns for referrals for 36–55-year-olds in both personal and professional peer group categories and that price sensitivity was not as significant for purchasers in this older age category as it was for those in the younger category. As a result, this disparity could be leveraged in the pricing strategy as well as the marketing strategy, by creating referral programs and incentives that were tailored to each market segment.

Figure 5.2 provides an overview of the inputs, tools, and outputs of the *Data Interpretation* process.

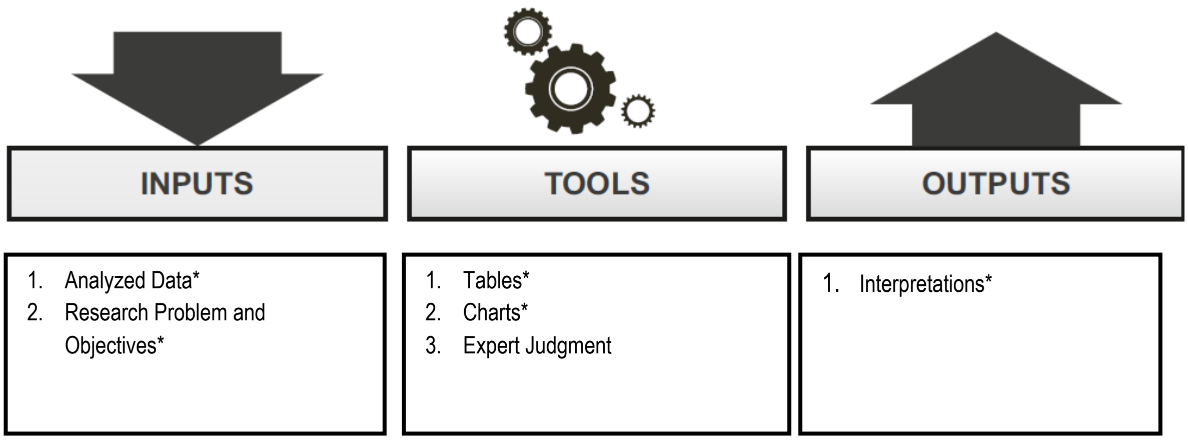


Figure 5-2: Data Interpretation—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

5.1.1 Inputs

5.1.1.1 Analyzed Data*

The analyzed data is an output of the *Data Analysis* process described in section 4.2. Properly analyzed data ensures that the interpretation process is efficient and that the reporting of the findings is meaningful. After the data collected have been processed and structured in a format that facilitates effective data analysis, the analyzed data is used to enable the team to identify patterns, draw conclusions, solve the research problem, and achieve the research objectives. The findings from the research analysis are compiled and reported to the marketing team and senior management and are ultimately used to inform marketing and business decisions.

5.1.1.2 Research Problem and Objectives*

The research problem and objectives provide a focused and definite direction to the data interpretation process. The research problem indicates the nature of the marketing decisions to be made, which in turn dictates or specifies the kind of information required to reach these decisions. The nature of the required information must always be specified and understood by the team responsible for solving the research problem. For example, from the research problem and objectives, researchers would know that they need to interpret data to measure consumer preferences if consumer preferences are key criteria to be considered by the decision makers when choosing among various alternative approaches. A comprehensive understanding of the research problem and objectives will ensure that the team in this example focuses on consumer preferences as a significant contributor to solving the research problem.

Example of Research Problem and Objectives:

- A restaurant chain is experiencing high volumes of traffic at several locations, resulting in long waits and some complaints from customers. The business knows that to accommodate the demand, it can either expand the restaurant's hours of operations or establish additional franchises to better serve the higher volume markets. It surveys the identified communities to determine specific preferences, patterns, and needs, focusing specifically on when consumers in each location seek the restaurant's services. Understanding the research problem and objectives will help the research team understand the focus of the project, namely determining the best approach to meet the high volume of restaurant goers.

  From the research, the team identifies two types of communities: those characterized by a proportionately high percentage of shift workers and commuters, for whom extended hours would best address the volume, and those where a high volume at peak hours would be better serviced by an additional franchise.

  Understanding that ultimately the team needs to determine how to best service the high demand of patrons in the selected markets helps the team focus on specific criteria from the research survey that eventually informs the strategy for each segment.

5.1.2 Tools

5.1.2.1 Tables*

Tables make reports easy to understand and information easily accessible. Clear labelling and the use of legends increase the effectiveness of tables, charts, and other graphics. As a general rule, all tables should contain the following:

- An identification number corresponding to the list of tables
- A title that conveys the content of the table, also corresponding to the list of tables
- Appropriate column labels and row labels
- The type of data shown
- The source of the table content

There are a number of ways to produce tables. When preparing a report, it is sometimes easiest to create a table manually. However, when complicated tables have to be produced, it is advisable to use spreadsheet software.

Example of Tables:

- Spreadsheet software such as Microsoft Excel and Google Spreadsheet are useful in organizing and representing large amount of complicated data.

5.1.2.2 Charts*

i. Bar Charts

Bar charts, as the name suggests, consist of a series of bars of equal thickness and whose lengths represent the value of items. Bar charts show changes in a dependent variable at discrete intervals of the independent variable. For example, bar charts can compare one year with another or one subset of the population with another subset.

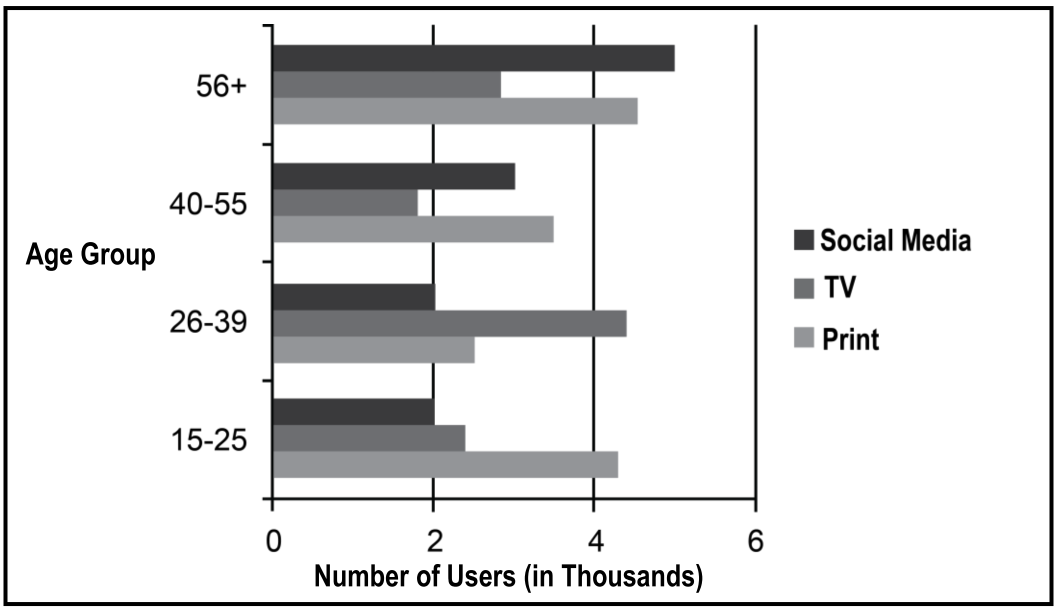


Figure 5-3: Sample Bar Chart

ii. Stratum Charts

A stratum chart is a two-dimensional graph with time along the horizontal axis and values of the items plotted along the vertical axis. Stratum charts are a variation of a line chart and have similar characteristics. They may show differences between variables more dramatically than line charts and may be displayed in three dimensions. They can be used as cumulative line charts and for showing comparisons of data based on various changing conditions. For example, stratum charts might be used to graph seasonal sales trends when no marketing communication efforts have been executed, sales trends after print and digital advertising campaigns have been implemented, and sales trends after advertising campaigns in combination with a coupon promotion have been implemented.

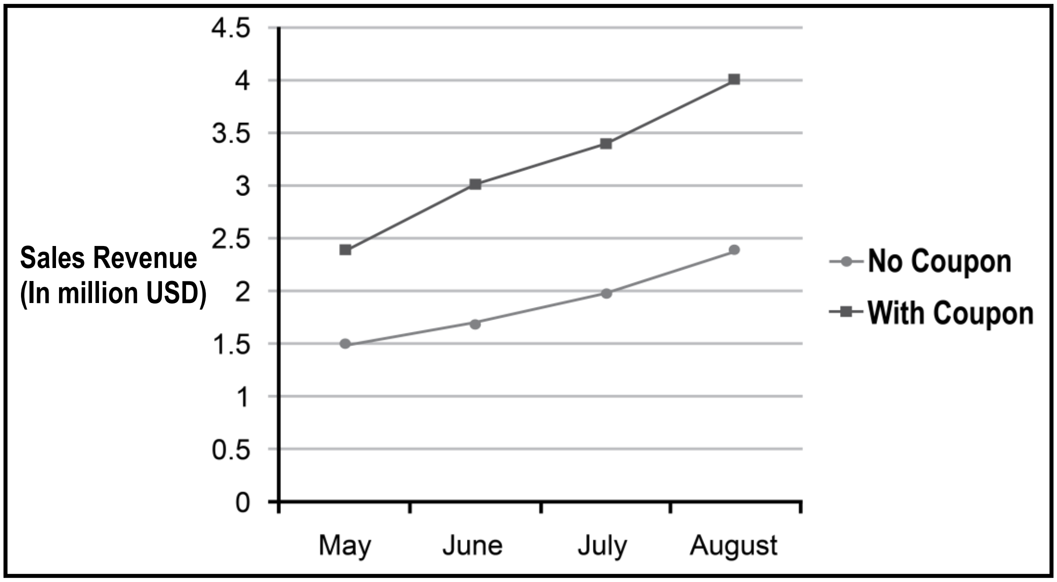


Figure 5-4: Sample Stratum Chart

iii. Pictograms

Pictograms are graphical displays used to present data. Pictograms are attractive and easy to understand, making them an effective method of communication. In pictograms, a pictorial symbol is used to clearly indicate the item that is being displayed. For example, for data relating to automobiles, a pictogram may use the symbol of a car. It should be noted that the symbols used in pictograms are not abstract presentations, such as a colored bar used in a bar chart; rather, they use easily identifiable visual cues.

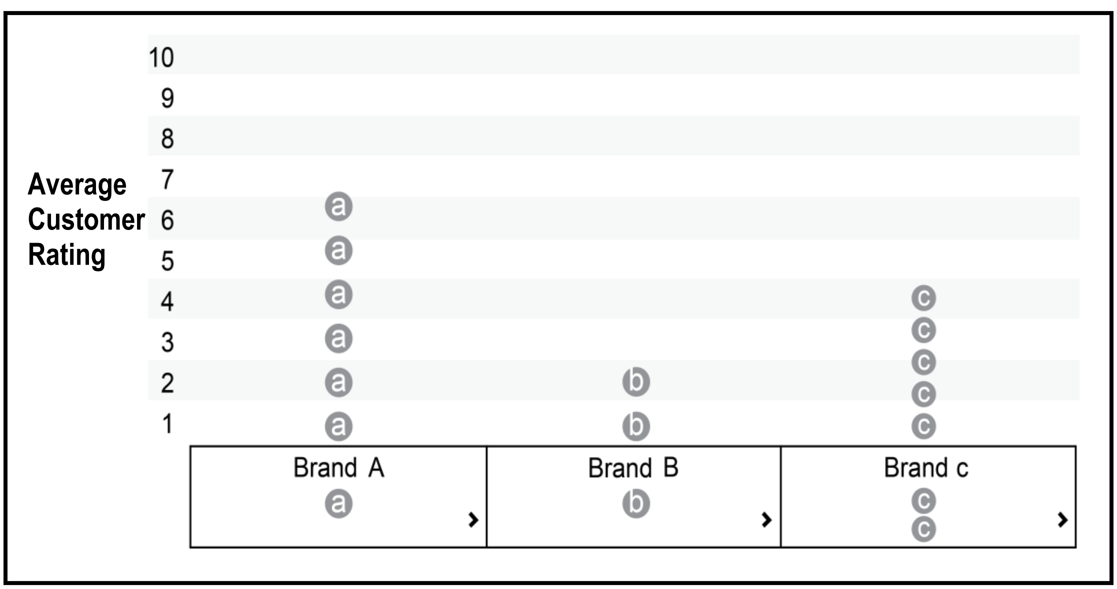


Figure 5-5: Sample Pictogram

iv. Cartograms

Cartograms are maps that are used to present statistical data that is geographically based. They are maps in which geographic entities such as states or countries are resized according to a particular parameter of interest, such as population, sales, market share, or electoral votes. Cartograms are very eye-catching. They are especially useful where geographic comparisons are significant and where approximate measures are acceptable. Figure 5.6 depicts a cartogram of the United States with states resized based on population.

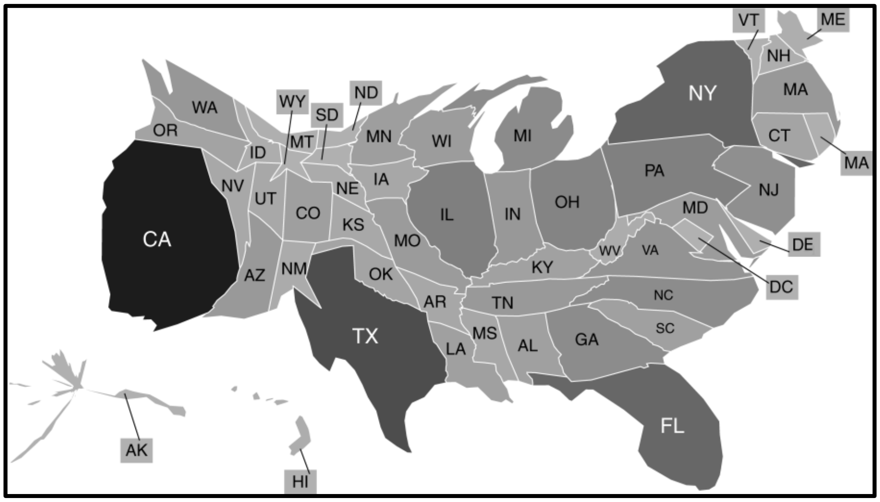


Figure 5-6: Sample Cartogram

5.1.2.3 Expert Judgment

The ability to appropriately interpret the data develops with experience. Inexperienced researchers can sometimes interpret data in a preferred way because of their comfort level with a given method. A researcher should try to seek the opinions of industry experts and research experts, who can provide valuable inputs in choosing the best way to interpret data within the given constraints.

5.1.3 Outputs

5.1.3.1 Interpretations*

Interpretations are key outputs from this process in which the analyzed data are converted into information that is both relevant to the research purpose and offers insights for the researcher to solve the research problem. With the help of interpretation tools, such as various charts, tables, and graphs deemed optimal for viewing and interpreting the data, the researcher draws conclusions that are used to solve the research problem and inform marketing decisions. Interpretations from this process will provide the researcher with an understanding of what has been observed during the research project, as well as a theoretical notion of any further research that needs to be conducted. Data analysis and interpretation are interlinked and are often performed simultaneously. The interpretation of the data is a key input for reporting the research findings.

5.2 Reporting

The results of a Marketing Research project must be effectively communicated to the sponsoring party. The consequences of incorrect reporting can be significant. Reporting errors have the potential not only to result in wasted resources but also to cause the making of poor decisions. Therefore, creating an accurate report with a compelling presentation of the facts to support or refute the business decision and research problem is imperative. Reporting is the final process of a Marketing Research project.

The results of a formal Marketing Research project are usually communicated to the management or sponsoring party with a formal written report and an optional oral presentation. The report and presentation are very important in terms of providing lasting documentation of the time and effort that went into the project. The report and presentation are the only part of a research project that decision makers usually see; therefore, their evaluation of the research project itself depends almost entirely on the quality of the report and its companion presentation.

The inputs, tools, and outputs of the process of *Reporting* are shown in Figure 5-7.

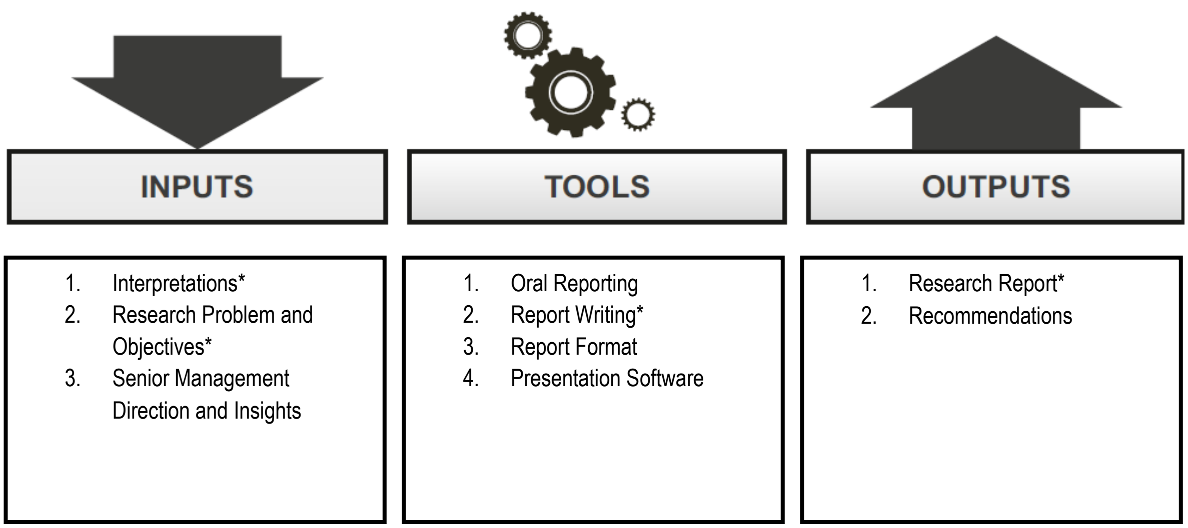


Figure 5-7: Reporting—Inputs, Tools, and Outputs

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

5.2.1 Inputs

5.2.1.1 Interpretations*

Interpretations are an output of the *Data Interpretation* process. Interpretation involves discerning or deriving meaning from the specific data that has been collected, processed, and analyzed. Data interpretation and data analysis are closely linked and often conducted simultaneously. Interpreting the data appropriately, accurately, and from the right perspective is important to ensure that sound marketing decisions are made based on the interpretations. The researcher should possess the skills and knowledge to interpret data with utmost care and objectivity. Any errors in data interpretation can nullify even the best research.

Interpretation should look at encapsulating the following points:

- gauging specifically what the gathered information means
- determining the implications for the client—how does this information impact the client?

The research report is written and the conclusions and recommendations are determined based on the interpreted data. Because of this, data interpretation is a major input for the reporting and writing of the research report. Data interpreted with errors may lead to unhelpful or even harmful recommendations and conclusions.

5.2.1.2 Research Problem and Objectives*

The purpose of conducting Marketing Research is to find key answers to management problems or research problems. The research should essentially derive a conclusion or recommendation for the research problem and objectives. While data analysis and interpretation provide statistical results, reporting translates these values into a language that is best understood by all interested parties of the research project.

The research problem and objectives is a key input for research reporting or report writing since the recommendations or conclusions should answer the main research objectives and problems. Understanding and considering this particular input is very important because the reporting has to be done in an unambiguous manner and chances of misinterpretation must be minimized.

5.2.1.3 Senior Management Direction and Insights

The researcher may seek some expert advice before or while writing the research report. Advice or feedback from senior management will help the researcher identify any deficiencies in the report regarding both the subject matter and the write-up.

Senior management direction and insights may help to answer a very important question. Does the report address the research objectives? Sometimes, the research problem and objectives is either overlooked or diluted toward the end of the research. This can be avoided by seeking advice from senior management, thus helping the researcher achieve accurate, efficient, and valuable research reporting.

5.2.2 Tools

5.2.2.1 Oral Reporting

An oral report is the presentation of research findings through verbal communication. The conclusions and recommendations of most research reports are presented orally as well as in writing. The purpose of an oral report is to communicate a project's most important findings and provide stakeholders an opportunity to ask questions. The oral reporting may be as simple as a short video conference with a manager at the client organization's location, or as formal as a presentation to the company board of directors.

The key to effective oral reporting is preparation. First, the researcher should know the audience so that the report's objective(s) can be clearly communicated. The objectives and approach of an oral presentation to senior management or a board of directors will be different than the objectives and approach used for a group of line managers. The presentation must be designed to address the interests and needs of the audience. For this purpose, the researcher should determine the backgrounds, interests, and involvement of the intended audience in the research project, as well as the extent to which they are likely to be affected by the project. The presentation should be rehearsed several times before it is given to management. When presenting an oral summary of a written report, all major points must be covered. At the same time, care should also be taken not to include excessive details to avoid loss of focus and the audience's interest. Delivering an hour-long formal speech when a ten-minute discussion is called for (or vice versa) will reflect poorly on both the presenter and the report. The presenter should ensure that within the time allotted, all major points and recommendations are covered. Also, the presenter should not use any research jargon that may confuse the audience. It's important that the presenter recognize that the audience may not have the same level of familiarity with certain terms and concepts that researchers are typically comfortable with. Thus, it is recommended that the presenter use simple language and familiar words during the oral presentation. The presenter should maintain eye contact with the audience and emphasize or repeat the main points. As well, it is advisable to identify possible questions that can be anticipated and have appropriate responses prepared.

The presenter may use visual aids such as charts, diagrams, slides, pictures, and so forth during the presentation to make the presentation more effective. Visual aids such as tables and graphs should be displayed with a variety of media. Slide shows or large pads of blank paper mounted on an easel enable the researcher to explain the findings. They are particularly useful in communicating answers to technical questions. The researcher may also distribute a handout containing important points, data, and charts for the audience to refer to.

Example of Oral Reporting:

- A municipal government hired NextReporting, a marketing research firm, to review user satisfaction with the public transportation system. After conducting interviews, phone surveys, and user observations, the research firm consolidated and organized its findings.

  NextReporting will present the results of these findings to the civic board in charge of the study. They will create both an oral presentation and a slideshow. Slideshows can be created using software such as using Microsoft Powerpoint or Adobe Acrobat, itemizing the top five concerns of the public, the top five items of excellence as identified by the public, and the top five recommendations for improvements. Presenting their findings in this fashion ensures that the associated report is understood and the key findings are communicated.

5.2.2.2 Report Writing*

A Marketing Research report is written for the marketing decision makers who will use the results of the research. While writing the report, the writer should keep the readers in mind. The researcher should consider the backgrounds, interests, and involvement of the intended readers in the project; the circumstances under which they will read the report; and the manner in which the report will be used. It is preferable to use descriptive explanations instead of technical marketing jargon. If using a technical term is essential, then a brief explanation of the term needs to be included in the appendix.

The clarity and the readability of any report are greatly enhanced by attention to matters of document structure and presentation style as well as content. The report should be written clearly and should be easy to read and comprehend. The flow and structure of the report should be logical. Proper headings should be given for each section so that the reader can easily navigate the document. Different people in an organization may have different levels of interest in the project and varying degrees of understanding of the subject. Structuring the content under different headings will save time, allowing readers to focus more intently on those portions that are most relevant to them.

The report should be written in professional language using an established style guide. The title page should be professionally designed; headings, subheadings, and the body of the content should be formatted uniformly. Attention should also be given to the following:

- blank space around key features/paragraphs
- use of underlining and italics
- font style and size in different sections of the document
- reinforcing key information in the text with tables, graphs, pictures, and maps
- the overall look and layout
- maintaining a consistent style
- infographics

The report should be true to its findings. Specifically, the report should accurately present the research design, results, and conclusions without modifying any content for a desired outcome.

Once written, the report needs to be checked thoroughly for completeness and errors. While the report should be complete and thorough, it should also be concise and to the point. Too much information can dilute the essence of the report. Anything unnecessary should be deleted. It is a good idea to have the report peer reviewed by someone else in the research organization. A person who is not involved in the research project can bring a fresh perspective and provide unbiased feedback.

Example of Report Writing:

- In the municipal government example, NextReporting will also deliver a detailed report on their findings. The report will reiterate the points covered in the oral presentation, and will summarize all comments and feedback provided by the surveyed population. Spreadsheets and graphs will be used to consolidate and organize the raw data. The intent of the report is to provide a comprehensive aggregation of all findings in the study that can be revisited, reviewed, and shared with other team members.

5.2.2.3 Report Format

The exact format of Marketing Research reports will vary from one researcher to another, from one research agency to another, and from one project to another, based on specific client requirements. While there is no standardized reporting format used in all research reports, the following outline provides the elements commonly included:

1. Title Page
2. Tables of Contents
3. Executive Summary and Recommendations
4. Problem Statement and Research Objectives
5. Research Methodology and Limitations
6. Data Analysis and Findings
7. Conclusions and Recommendations
8. Appendices and Supporting Documentation

1. **Title Page**—The title page includes the title or subject of the research study, who it was prepared for, information about the researcher or organization conducting the research, and the date of release. The title page may also bear the logo of the research organization, if applicable. If the report has been revised, this revision should be noted on the title page.

2. **Table of Contents**—The table of contents should list the major headings and subheadings in the report with the appropriate page numbers. There may also be lists of tables, graphs, appendices, and exhibits. If the entire report is only ten to twelve pages in length, then a table of contents is not necessary. For longer reports, it is an essential element.

3. **Executive Summary and Recommendations—**The executive summary is an extremely important part of the report. It presents the highlights of the research report in a straightforward and precise manner. Usually this is the only section that is read by the senior decision makers in the client organization. This summary should be comprehensive enough to enable the reader to understand the essence of the findings without reading through the detailed report. The summary should briefly describe the research problem, approach, and the research design. It should include a section that highlights the major results, conclusions, and recommendations.

4. **Problem Definition and Research Objectives**—The section elaborates the situation the management is facing and the background of the problem. This is the place to list the management problem, the research purpose (including decision alternatives and criteria), and the research objectives and hypotheses.

5. **Research Methodology and Limitations**—This section should summarize the approach and research methodology adopted to address the problem. Justification should be provided for adopting the specific research approach and method. The section should detail the methods undertaken in the data collection from secondary and primary sources. Care should be taken not to include too much detail because executives are typically not interested in the small details of how the research was conducted. These topics should be presented in a nontechnical, easy-to-understand manner. The technical details should be included in an appendix. Any limitations relating to a skewed sample, low response rate, or timing of the study should be highlighted because these may have a bearing on the study's findings and conclusions. In general, this section should not be longer than four to five pages, with further technical details being provided in an appendix.

6. **Data Analysis and Findings**—This section describes how data analysis was performed, the tools and techniques that were used, and the rationale for the data analysis techniques and approach. This section also typically includes the results or findings of the project; however, some authors prefer to write the findings in a separate section. The findings section is the longest section in a research report and should be written in a logical and organized manner. This section describes the findings in detail with supporting graphs and charts included as both references and sources of support for the statements that are made in the narrative of this section. The organization of the findings should be geared directly to the components of the Marketing Research problem and the information needs that were identified. The nature of the information needs and characteristics of the recipients of the report ultimately determine the best way to present results. The major findings should be arranged with each corresponding research objective so that the reader is taken, step-by-step, though a progression of all of the findings of the study. This approach also ensures that all the objectives are covered in the write-up of the findings.

7. **Conclusions and Recommendations**—The researcher's role does not end in presenting the findings, but extends to interpreting the findings in light of the research problems. Identifying the main implications for the client is also the researcher's responsibility. The researcher should present a set of conclusions and recommendations for managerial action. While the final decision belongs to the decision maker, it is the job of the researcher to recommend a course of action based on the understanding garnered from involvement in the project. The conclusions should be concise, highlighting key findings and implications as well as providing justification for the conclusions and recommendations.

8. **Appendices and Supporting Documentation**—Additional information that might be useful to the readers of the report should be provided at the end of the report in the appendix. The questionnaires used in the study, more detailed secondary data, detailed statistical analysis, or any other type of complex or specialized data not directly relevant to the research objectives can be included in this section. Any material that may be interesting to the client but would interrupt the flow and coherence of the document if it is included within the main body of the report should be placed in the appendices. This section may also include lists of contacts, references used, and further sources of reference.

5.2.2.4 Presentation Software

There are various software packages that can be used in writing a research report or presenting it orally. Microsoft Office and Adobe Acrobat are examples of widely used software packages for this purpose. There are many other online reporting programs that improve the efficiency of report writing and presentations to decision makers. There are several popular software tools that allow users to create different types of tables, charts, and graphics to present the content in a more effective manner.

With presentation software, professional slideshows can easily be created with graphics and animation for oral presentations. Most presentation software packages also have features to enhance the dramatic presentation of ideas. Most of the presentation programs also enable the user to create master design templates, providing uniformity to the entire presentation.

Example of Presentation Software:

- Microsoft Office and Adobe Acrobat are examples of widely used software packages for report writing and presentation.
- Again, considering the municipal government research project, NextReporting will prepare a presentation for the client. NextReporting has extensive experience providing marketing studies. The company has a PowerPoint presentation template that is modified for the specifics of each study. This helps ensure that all the desired aspects of each individual study are communicated to the customer. The company also delivers the PowerPoint slides to the client with the associated written report, allowing the client to share the presentation with stakeholders who are not present at the oral presentation.

5.2.3 Outputs

5.2.3.1 Research Report*

The research report should be written for specific readers or a target audience. Usually, marketing managers are the readers who use the results. The research report can also be written to cater to the needs of several different audiences that are all interested in the project. A research report is considered to be efficient only when specific readers understand the report and accept the solution.

The presentation and appearance of the report is very important. The report should be professionally produced for a specific audience using various tools such as presentation software, oral reports, and printed documents.

The research objectives should provide a guide for the report writing process. The report should accurately present the findings, methodology used, results, and conclusions of the Marketing Research. The findings and conclusions should never be manipulated to adapt to management’s expectations. The readers should get an honest and complete disclosure of the research procedures and results after reading the research report.

5.2.3.2 Recommendations

The research report should always address the research problem in order to arrive at major conclusions. Based on the research results and conclusions, the researcher may make recommendations to senior management. Simply reporting or presenting a summary of the statistical results is typically not sufficient; however, in some cases, the researcher is not asked to make any recommendations. This is mainly because the research may be in a particular area with which the researcher has insufficient familiarity to make recommendations. The researcher may be providing only a segment of the information that is required for the client to make key decisions or choose a course of action. However, when the researcher in fact provides recommendations, they should be reasonable, concrete, and actionable and should provide insights to inform managerial decision-making.

Example of Recommendations:

- In the case of the user-satisfaction survey on public transit, the investigated research question was, “What are the primary areas of dissatisfaction with the service being offered?” Based on the findings of NextReporting, the board in charge of the initiative has recommended increasing the frequency of bus service on some of the major routes and communicating this to the general public. While the marketing research was performed by NextReporting, the ultimate decision on how to proceed is the responsibility of the civic board.

APPENDIX A.1: EVOLUTION OF SALES AND MARKETING

Since Sales and Marketing has evolved significantly over time, it is important to present a high-level overview of the history of the subject in order to understand and appreciate its relevance in the world today.

Figure A.1-1 depicts the timeline for the Evolution of Sales and Marketing.

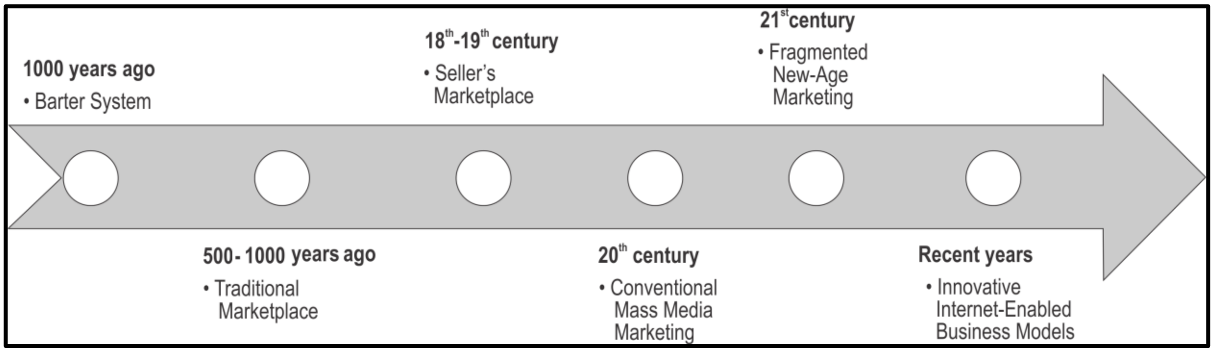


Figure A.1-1: Evolution of Sales and Marketing Timeline

A.1.1 Barter System

More than a thousand years ago, when coins and other forms of money were not yet popular, the typical and most common way people procured their products or services was through the barter system—the direct exchange of goods or services without the use of money. For example, a farmer might exchange some of his harvest with a carpenter for some wooden furniture. Sales and Marketing with the barter system is dependent on having access to the appropriate persons with whom things can be exchanged for mutual value to both parties.

Barter continues to be used today—people and countries still exchange some goods and services without the use of money. The barter system may replace money in times of monetary crisis, when the usual exchange currency is unavailable, or when currency is unstable (e.g., due to high inflation).

Examples of Barter System:

- A tradesperson, such as a carpenter or electrician, operating his or her own business, might provide services free of charge to his or her accountant in exchange for professional accounting services.
- Today, many websites provide a space for individuals to offer goods for bartering purposes. An individual planning to relocate to a crowded downtown location might offer his or her car in exchange for a more appropriate vehicle for downtown transportation, such as a motorcycle or a scooter.

A.1.2 Traditional Marketplace

Five hundred to thousand years ago, coins and other forms of money started becoming popular as a medium of exchange between people. This led to the creation of the traditional marketplace where producers, such as farmers, craftsmen, and carpenters create products, stay in their shop with their wares, and shout out to a crowd of potential customers in the marketplace in order to promote and sell their products.

Traditional marketplaces are usually small markets where price negotiations and other decisions related to sales are made quickly—often by one or two persons. There may be significant flexibility regarding discounts and additional product benefits. The focus is more on short-term gains and less on long-term transactions and relationships. There is negligible branding and advertising; rather the objective is to sell what has been produced.

Example of Traditional Marketplace:

- The traditional marketplace is still in use today, in some cases, under unique labels such as the bazaars of Turkey, the haats of India, the floating markets in Thailand, the wet markets in Hong Kong, the flea markets in Germany, the souks of the Middle East, the farmers' markets in the US, and the tianguis of Mexico.

A.1.3 Seller's Marketplace

The Industrial Revolution in the eighteenth and nineteenth centuries marked a shift to mass production in factories (e.g., textile manufacturing). During this time, transportation infrastructure improved significantly with inventions such as the steam engine and more efficient ships. The banking system was further developed and the exchange of money became easier. Communication was also substantially improved through the development of the postal system and the use of telegraphs. Furthermore, goods were produced more efficiently and economically in factories and could be sold to a wider market. This created the seller's marketplace.

The main objective of the seller's marketplace is to establish a supply chain to procure products and then establish a distribution channel to sell the products to a wide variety of customers, often referred to as "mass marketing." Emphasis on branding and advertising is minimal in a seller's marketplace.

Examples of Seller's Marketplace:

- The seller's marketplace continues to be used today in some countries where agricultural produce is often procured by the government. The government in turn manages the distribution of the produce to the different markets.
- The seller's marketplace is also prevalent in industries where the government controls the competition of private companies, for example, the distribution of petroleum products, or licenses that allow only a select few companies to manufacture a particular product in the country).
- If a natural disaster or unfavorable weather conditions caused widespread crop failure, a seller who had stores of that particular crop would be able to capitalize on a seller's marketplace because there would be more buyers than available product.

A.1.4 Conventional Mass Media Marketing

In the twentieth century, as the number of manufacturers or industries for specific products grew, consumers had the option to buy from multiple manufacturers. Unlike a seller's marketplace where sellers have the advantage over customers, mass media marketing features multiple manufacturers, thus shifting the balance of power in favor of consumers. Manufacturers created differentiated perceptions for their products by developing brands or names for their specific products or services with a specific message or positioning. They also began advertising their products or brands for a wider reach.

Primary channels used for mass media marketing are print advertising (newspapers, magazines, inserts, or run of paper), mass mailers (flyers, postcards), television (network, cable, or syndication), radio (national, local, satellite, or podcast), and outdoor advertising (billboards, bus shelters, stadiums). The viable channels for conventional mass media marketing may be restricted in some instances (i.e., some channels may be cost prohibitive or simply unavailable in some markets); however, a company can reach a wide segment of consumers using one or more channels effectively. For example, a business may choose to use only newspaper advertising and mass mailers to advertise the launch of a new business. It is also important to note that identifying the revenue generated from mass media marketing spend can help assess the success or failure of specific mass media marketing campaigns.

The objective of conventional mass media marketing is for organizations to create strong brands and differentiated brand perceptions so that consumers will desire and purchase their products rather than those available from competitors. Thus, mass media marketing usually uses cumulative repetition over time to influence consumer attitudes and purchase actions. Mass media marketing also involves creating distribution channels and appropriate pricing and positioning strategies to ensure that desirable products are available to customers at specific price points.

Example of Conventional Mass Media Marketing:

- Conventional mass media marketing continues to be used today, particularly by companies with established brands with relatively high marketing budgets and a broad target market. Companies such as PepsiCo, Coca-Cola, Procter & Gamble, Unilever, McDonald's, and Walmart, continue to primarily use mass media marketing for marketing their products and brands. In recent years, some of these companies have decreased their budgets for conventional mass media marketing, and have in turn increased allocations towards fragmented new-age marketing and/or innovative Internet-enabled business models. One of the key drivers for this change is the fact that consumers generally spend significantly more time online (i.e., using computers, tablets, and mobile phones) than they used to, so targeting them through conventional mass media marketing would be sub-optimal.

A.1.5 Fragmented New-Age Marketing

This is described in Section 1.3.2

A.1.6 Innovative Internet-Enabled Business Models

This is described in Section 1.3.3

A.1.7 Sales and Marketing as a Continuum

This is described in Section 1.3.4

APPENDIX A.2: MARKETING STRATEGY OVERVIEW

All successful products or brands need well-planned marketing strategies in place to ensure that they satisfy the goals set by the Corporate Marketing Strategy or Business Unit/Geographic Strategies. Marketing Strategy is one of the most crucial Aspects of Sales and Marketing. It defines a product or brand's unique value proposition, target markets, and strategies to connect with defined audiences. It also specifies the overall pricing and distribution strategies of the product or brand, and outlines the objectives, metrics, and budgets for all its marketing activities. So, the Marketing Strategy includes a set of outputs from the eleven marketing processes mentioned below, which provides an overall direction to the product or brand for all its marketing initiatives. Figure A.2-1 provides an overview of the important processes and outputs of Marketing Strategy.

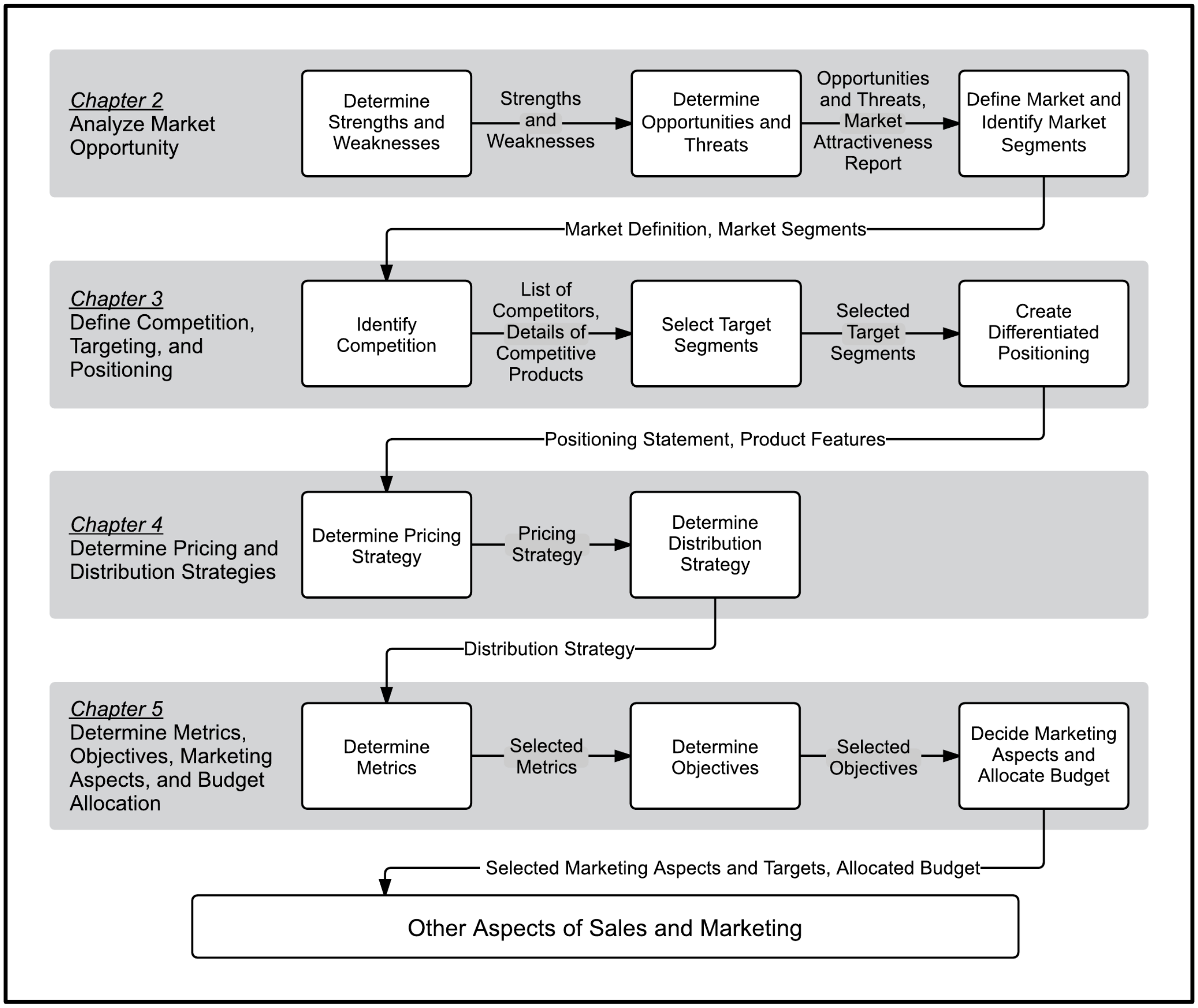


Figure A.2-1: Summary of the Marketing Strategy Processes

A.2.1 Analyze Market Opportunity

Analyze Market Opportunity is the second chapter of this book on Marketing Strategy. This chapter discusses concepts related to analyzing the external environment and evaluating the internal capabilities of a company. An analysis of market opportunities is important because businesses operate in dynamic and constantly evolving environments, so understanding the changing landscape and trends that are impacting the business helps in developing an effective Marketing Strategy. This chapter explains the crucial points while analyzing market opportunity, such as defining the market within which a company intends to operate, and segmenting the market to identify key customers for the company's product portfolio.

Each process associated with Analyze Market Opportunity is discussed in detail. The processes are: *Determine Strengths and Weaknesses*, *Determine Opportunities and Threats*, and *Define Market and Identify Market Segments*. Each process is explained through its associated inputs, tools, and outputs. Analyze Market Opportunity helps an organization understand what it can deliver so that it can fulfill customer needs.

Figure A.2-2 provides an overview of the processes pertaining to Analyze Market Opportunity, which is discussed in detail in chapter 2 of the *SMstudy® Guide* book on Marketing Strategy (MS).

2.1 Determine Strengths and Weaknesses

INPUTS
1. Senior Management Direction and Insights*
2. Organizational Capabilities*
3. Assumptions and Constraints*
4. Existing Marketing Research Reports

TOOLS
1. Meetings and Discussions*
2. Product Portfolio Analysis*
3. BCG Growth-Share Matrix
4. Value Chain Analysis
5. Marketing Research

OUTPUTS
1. Strengths and Weaknesses*
2. Marketing Research Reports

2.2 Determine Opportunities and Threats

INPUTS
1. Senior Management Direction and Insights*
2. Existing Marketing Research Reports
3. Generic Reports

TOOLS
1. Meetings and Discussions*
2. Porter's Five Forces for Industry Attractiveness
3. Market Analysis
4. Marketing Research
5. PESTEL Analysis*

OUTPUTS
1. Opportunities and Threats*
2. Market Attractiveness Report*
3. Marketing Research Reports

2.3 Define Market and Identify Market Segments

INPUTS
1. Strengths and Weaknesses*
2. Opportunities and Threats*
3. Market Attractiveness Report*
4. Existing Goals
5. Existing Marketing Research Reports

TOOLS
1. Meetings and Discussions*
2. Demographic Segmentation
3. Psychographic Segmentation
4. Behavioral Segmentation
5. Company Characteristics Based Segmentation
6. Evaluation of Future Market Scenarios*

OUTPUTS
1. Market Definition*
2. Market Segments*

Figure A.2-2: Analyze Market Opportunity

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

A.2.2 Define Competition, Targeting, and Positioning

The third chapter of Marketing Strategy is Define Competition, Targeting, and Positioning. This chapter first covers the activities of identifying the competition, understanding industry trends, and creating future competitive scenarios that help in selecting target market segments. It then looks at creating a differentiated positioning statement for the company's products or services for the selected target segments. Competitive positioning tools help define how a company can differentiate its product offerings to create value in the market by fully understanding its target segments and the competitive landscape.

There are three processes outlined that help an organization understand market competition, target appropriate market segments, and define product features that help create a differentiated positioning statement for the products or services of the company. The processes discussed in this chapter are *Identify Competition*, *Select Target Segments*, and *Create Differentiated Positioning*.

Figure A.2-3 provides an overview of the processes pertaining to Define Competition, Targeting, and Positioning, which is discussed in detail in Chapter 3.

3.1 Identify Competition

INPUTS
1. Senior Management Direction and Insights*
2. Market Segments*
3. Existing Marketing Research Reports
4. Information Published by Competitors

TOOLS
1. Competitor Selection Criteria*
2. Future Competitive Analysis*
3. Marketing Research
4. Meetings and Discussions*

OUTPUTS
1. List of Competitors*
2. Details of Competitive Products*
3. Industry Benchmarks and KPIs
4. Future Competitive Scenarios
5. Marketing Research Reports

3.2 Select Target Segments

INPUTS
1. Strengths and Weaknesses*
2. Opportunities and Threats*
3. Market Segments*
4. List of Competitors*
5. Details of Competitive Products*
6. Industry Benchmarks and KPIs
7. Future Competitive Scenarios
8. Existing Marketing Research Reports

TOOLS
1. Market Segment Attractiveness Matrix*
2. Undifferentiated Strategy
3. Focused or Concentrated Strategy
4. Differentiated Strategy
5. Market-Product Grid
6. Customer Personas*
7. Meetings and Discussions*
8. Marketing Research

OUTPUTS
1. Selected Target Segments*
2. Marketing Research Reports

3.3 Create Differentiated Positioning

INPUTS
1. Selected Target Segments*
2. Strengths and Weaknesses*
3. Opportunities and Threats*
4. List of Competitors
5. Details of Competitive Products
6. Industry Benchmarks and KPIs
7. Marketing Research Reports
8. Customer Feedback

TOOLS
1. Selecting Points of Parity and Differentiation*
2. Perceptual Maps*
3. Product Categories
4. Meetings and Discussions*
5. Marketing Research

OUTPUTS
1. Positioning Statement*
2. Product Features*
3. Updated Corporate Strategy

Figure A.2-3: Define Competition, Targeting, and Positioning Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

A.2.3 Determine Pricing and Distribution Strategies

Determine Pricing and Distribution Strategies is the fourth chapter of Marketing Strategy. A Pricing Strategy properly prices a company's product or service so that the company can sustain profitability while maintaining or growing its market share. This involves assessing the value of the company's product based on its features; analyzing the pricing and features of competitive products in the market; analyzing the consumer mindset, which takes into account demand and price expectations for the product; and considering anticipated unit sales and costs and, in turn, profitability. A Distribution Strategy defines how a company moves a product from creation to consumption in a cost-efficient manner while focusing on end users' needs. The concepts in the Distribution Strategy are important because understanding and addressing the needs of the entire distribution channel external to the company ensures that a product or service is delivered and sold to customers in the best possible way.

This chapter explains the two processes associated with Determine Pricing and Distribution Strategies. These are *Determine Pricing Strategy* and *Determine Distribution Strategy*. The Pricing Strategy is determined for the various products or services of a company. The end objective is sustainable profitability while growing or maintaining a healthy market share. The Distribution Strategy ensures the most efficient delivery of a company's products or services to the customer and that the selected strategy is based on the company's assessment of several alternative distribution channels. These processes are explained with the help of their associated inputs, tools, and outputs.

Figure A.2-4 provides an overview of the processes pertaining to Determine Pricing and Distribution Strategies, which is discussed in detail in Chapter 4.

4.1 Determine Pricing Strategy

INPUTS
1. Market Attractiveness Report*
2. Past Revenues and Cost Data*
3. Positioning Statement*
4. Marketing Research Reports
5. Purchase Timing
6. Existing Warranties
7. Opportunities and Threats
8. Details of Competitive Products*
9. Product Features

TOOLS
1. Marketing Research
2. Price Elasticity of Demand*
3. Total Cost of Ownership*
4. Economies of Scale
5. Experience/Learning Curve
6. Target Costing
7. Break-even Analysis*
8. Perceived Value Pricing
9. Value Pricing
10. Going Rate Pricing/Neutral Pricing
11. Auctions
12. Gain and Risk Sharing Pricing
13. Variations among Product Geographies
14. Discounts and Offers*

OUTPUTS
1. Pricing Strategy*
2. Marketing Research Reports

4.2 Determine Distribution Strategy

INPUTS
1. Positioning Statement*
2. Product Features*
3. Pricing Strategy*
4. Opportunities and Threats
5. Marketing Research Reports
6. Performance of Existing Channel Members

TOOLS
1. Demand Chain Planning
2. Value Network Analysis
3. Evaluation of Distribution Channel Models*
4. Marketing Research
5. Meetings and Discussions*

OUTPUTS
1. Distribution Strategy*
2. Total Cost of Distribution
3. Updated Pricing Strategy
4. Marketing Research Reports

Figure A.2-4: Determine Pricing and Distribution Strategies Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

A.2.4 Determine Metrics, Objectives, Marketing Aspects and Budget Allocation

This chapter deals with determining the various metrics and objectives used for Sales and Marketing such as reach, brand perception, product availability, sales, and profitability. It also provides an overview of various Sales and Marketing Aspects, including Marketing Research, Digital Marketing, Corporate Sales, Branding and Advertising, and Retail Marketing, and presents a framework for allocating targets and budget for each of these Aspects.

This chapter is broken down into three processes explained with the help of their corresponding inputs, tools, and outputs. The processes are *Determine Metrics*, *Determine Objectives*, and *Decide Marketing Aspects and Allocate Budget*. In the first process, *Determine Metrics*, various Sales and Marketing metrics such as reach, brand perception, product availability, sales, and profitability are determined. These metrics help to measure the success or failure of the Marketing Strategy. In *Determine Objectives*, attainable, quantifiable and time-based objectives are determined for all of the metrics selected in the previous process. In the final process, the Sales and Marketing teams select the Marketing Aspects that will help the company reach its overall Sales and Marketing objectives. Subsequently, specific objectives are determined for each Marketing Aspect, and a marketing budget is also allocated for each.

Figure A.2-5 provides an overview of the processes pertaining to Determine Metrics, Objectives, and Budget Allocation, which is discussed in detail in Chapter 5.

5.1 Determine Metrics

INPUTS
1. Positioning Statement*
2. Pricing Strategy*
3. Distribution Strategy*
4. Industry Benchmarks and KPIs

TOOLS
1. Meetings and Discussions*
2. Product Life-cycle Analysis
3. SMART Framework*
4. Customer Reach Metrics*
5. Brand Perception Metrics*
6. Product Availability Metrics*
7. Sales and Profitability Metrics*

OUTPUTS
1. Selected Metrics*

5.2 Determine Objectives

INPUTS
1. Selected Metrics*
2. Senior Management Direction and Insights*
3. Past Performance Data
4. Market Attractiveness Report
5. Future Competition Scenarios

TOOLS
1. Meetings and Discussions*
2. Trend Analysis
3. SMART Framework*

OUTPUTS
1. Selected Objectives*

5.3 Decide Marketing Aspects and Allocate Budget

INPUTS
1. Selected Objectives*
2. Senior Management Direction and Insights*
3. Available Budget
4. Past Performance Data
5. Details of Competitive Products

TOOLS
1. Meetings and Discussions*
2. ROI Comparison of Aspects*
3. Analysis of Competition's Marketing Aspects

OUTPUTS
1. Selected Marketing Aspects and Targets*
2. Allocated Budget*
3. Updated Objectives

Figure A.2-5: Determine Metrics, Objectives, and Budget Allocation Overview

Note: An asterisk (*) denotes a highly recommended input, tool, or output for the corresponding process.

APPENDIX A.3: STATISTICS

A.3.1 Data

Statistics is the study of the collection, measurement, analysis, and interpretation of numerical data. A list of numerical data can be described by various statistical measures, in particular, the mean (average), median, mode, range, and standard deviation. Each of these measures and their properties will be discussed in this section and demonstrated with examples.

Data are pieces of information pertaining to a particular subject(s) of study. Data are collected by taking measurements of selected attributes or features of the subject(s) under study. Data can be numerical or non-numerical. In most marketing research situations, marketing professionals deal with numerical data. For example, the data for a marketing research project might include customer online spend patterns for a retail sporting goods store.

Here the data to be collected are the spend patterns for each customer. The subjects are the customers. Now, what is being measured? In marketing research, the researcher is measuring a specific feature or characteristic of a subject. This is known as a variable in statistical terminology. Here one of the variables is how much money each customer spends per online purchase. How is this measured? Because customers are required to create an account before they purchase, their spend patterns and demographic details are stored in a database and can be tracked, measured, and graphed. These numerical values constitute the data.

A.3.1.1 Types of Numerical Data

Numerical data are broadly classified as discrete data and continuous data. Data generated by counting is known as discrete data whereas data generated by using some measurement device or physical instrument is known as continuous data. The customer spending collected in the previous scenario is an example of discrete data because the data is collected by counting the spend per customer. The differences between these two types of numerical data are as follows:

Discrete Data:

- Can be produced by counts
- Can be classified/categorized into certain groups
- Examples: number of visitors to a website, number of people who visit a retail storefront, number of inbound calls, number of times customers repurchase products, and binary classifications (e.g., good/bad, yes/no, pass/fail)

Continuous Data:

- Is measured on a physical instrument
- Can be a fraction of a whole number
- Can be further broken down into smaller parts
- Examples: length, height, width, volume, time, temperature, viscosity, and velocity

A.3.2 Statistical Measurement of Data

The various statistical measures of data used in descriptive statistics are as follows:

- Measures of central tendency such as mean, median, and mode give an indication of the center/location of the data.
- Measures of dispersion such as range and standard deviation give an indication of the spread of the data from the central location.
- Measures of symmetry such as skewness and kurtosis. These give an indication of the shape of the distribution of data i.e. whether the data is distributed symmetrically (evenly on either side of the central location) or whether a greater percentage of data is concentrated either to the right or left of the central location.

A.3.2.1 Mean

The mean, also known as the arithmetic mean or average, is defined as the sum of all the members in a set divided by the total number of members.

Example of Mean Calculation:

- The following numbers represent the daily count of unique visitors to a web page over a ten-day period: 83, 67, 90, 78, 81, 59, 74, 60, 84, 77

  Calculating the average or mean unique visitors can provide a marketer with a benchmark against which to compare data gathered after a marketing campaign in order to determine the success of the marketing activity.

  Calculation:
  First, calculate the sum of all the values, (i.e., the total number of unique visitors over the ten-day period), which = 83 + 67 + 90 + 78 + 81 + 59 + 74 + 60 + 84 + 77 = 753

  Because there are 10 days of data being examined, the total number of elements in the set is 10. Therefore,
  the average daily unique visitors $= \frac{753}{10} = 75.3$

Change in Mean Due to Other Factors

Often it is important to examine the change in *mean* due to the addition of one or more elements to the set.

Example of Change in Mean:

- Considering the previous example, we can add another day of data, with a total of 41 unique site visitors, and calculate the new average daily site visitors.

  Calculation:

  New average = $\frac{\text{Total sum}}{\text{Total number of members}} = \frac{753+41}{10+1} = \frac{794}{11} = 72.2$

  It is clear that the new average of 72.2 is different and less than the previous average of 75.3. This reduction is due to the addition of the data from a day on which the site gained significantly fewer unique site visitors (41) than were gained on the other 10 days. Therefore, the mean can change (either increase or decrease) due to the addition or deletion of one or more elements to the set. The extent of change depends on the value of the element(s) added or deleted.

In some instances, the *mean* can remain unchanged following the addition or deletion of one or more members to the set. This scenario occurs in the following situations:

- If the initial mean is equal to the value of the member(s) added or deleted
- If all the members of the set have equal values as that of the added or deleted member(s)

Some Properties of a Mean:

- If we add (or subtract) a constant value to each member of a set, the new mean can be obtained by adding (or subtracting) the constant to the initial mean.
- If we multiply (or divide) a constant value to each member of a set, then the new mean can be obtained by multiplying (or dividing) the constant to the initial mean.

A.3.2.2 Median

The median is the middle value of a set whose values are arranged in ascending order. It divides the set into two equal halves. Values in one half are less than or equal to the median, and values in the other half are greater than or equal to the median.

How to Calculate the Median:

- Arrange the values of a set in ascending order

- If the number of values (n) is odd, then the median is the middle, or $\left(\frac{n+1}{2}\right)^{th}$ value
- If the number of terms (n) is even, then the median is the average of the two middle, or $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ values

Example of Median:

- Using median rather than mean as a way of examining certain trends or ratings is sometimes useful when there is potential for data to be skewed by values in a set that are abnormally high or low, resulting in a mean that could be misinterpreted.

  The following numbers represent the number of unit sales per day for a specific inventory item: 40, 29, 36, 2, 8, 18, 0, and 35

  Calculation:

  1. Arrange the numbers in ascending order: 0, 2, 8, 18, 29, 35, 36, 40
  2. Determine the number of values in the set: 8
  3. Identify whether the number is even or odd.
  4. Since 8 is even, the median will be the average of the two middle values, i.e. $\frac{8}{2}$ = the 4^{th} and $\left(\frac{8}{2}+1\right)$ = 5^{th} elements
  5. The 4^{th} and 5^{th} elements of the set of numbers (arranged in ascending order) are 18 and 29
  6. So, the median = $\frac{18+29}{2} = \frac{47}{2} = 23.5$

  In this example the mean of the same set of numbers is 21, but the mean is skewed given the fact that one of the values is 0. There may be a number of reasons for the 0 value in the data, including an out of stock problem. Using the median rather than the mean in this situation would provide a more accurate indication of daily unit sales of the particular inventory item.

A.3.2.3 Mode

The mode of a set is the element with the highest *frequency* of occurrence. The mode is determined by the number of times an element repeats itself in a set. A set may have *more than one mode* if several values have the same frequency of occurrence.

How to Calculate the Mode:

- Write all the distinct values of the set in a column
- Count the number of occurrences for each value and write them accordingly in the adjacent column corresponding to each value

- Identify the value with the highest frequency; that will be the mode

Examples of Mode:

- The following numbers represent a set of customer ratings from 0–9 for customer service at a restaurant: 3, 4, 4, 0, 0, 2, 3, 4, 3, 6, 4, 5, 7, 6, 7, 0, 2

  Calculating the mode of the ratings provides the marketer with the most often selected rating from the options provided.

  Calculation:

  1. Identify the unique elements in the list.
  2. Determine the frequency or the number of times each element is repeated in the list and present it in tabular form (called the *Frequency Table*).
  3. Identify the elements with the highest frequency.

Elements	Frequency
0	3
2	2
3	3
4	4
5	1
6	2
7	2

  It is clear from the table above that a rating of 4, with a frequency of 4 has the highest frequency. Thus, the mode of the given set is 4.

- Mode is a useful measure when considering qualitative data, such as customer's preferred flavors or colors. A market research survey might ask consumers to indicate their preferred color of car, the results would be tabulated as follows:

Elements	Frequency
black	23
blue	16
silver	25
red	18
tan	7
white	31
grey	11

In this sample, the mode is white.

A.3.2.4 Interrelation between Mean, Median, and Mode

The mean is dependent on the value of each element in a set, whereas median and mode both depend on the number of elements in a set rather than on their individual values. The mean is equal to the median of a dataset if the values are in arithmetic progression (AP) (i.e., the values are evenly spaced: the difference between any two consecutive values of the set is constant). In this case, the mean or median can be obtained by taking an average of the extreme values or average of the two values equidistant from the extreme values. It is not necessary to calculate the mean of the entire set.

A median is not affected greatly due to the addition (or deletion) of an element with a very low, or a very high value. However, a mean can vary considerably depending on the value of the added (or deleted) element.

A.3.2.5 Range

Mean, median, and mode describe the central tendency of a data set, whereas the range describes the spread of a data set. The range is calculated as the difference between the highest value and the lowest value of the set, also known as the extreme values. The range cannot be negative. It will be zero if there is only one value in the set, or if all the values are the same. Although range does measure the dispersion of a data set, it is considered a poor measure because it takes only two extreme values (the highest and the lowest) of the set, and thus does not completely describe the data set. A better measure of dispersion is standard deviation (SD).

Example of Range:

- The following numbers represent the page views per day of a particular web page: 45, 28, 39, 2, 8, 11, 0, and 38.

  To determine the range of the set of numbers, the researcher must take the following steps:
  1. Identify the highest and lowest elements in the set
  2. Range = Highest element – Lowest element

  Here, the highest element = 45 and the lowest element = 0
  Thus, the range = 45 – 0 = 45

A.3.2.6 Variance and Standard Deviation

1. Variance

Variance (*V*) is the measure of the spread of numbers from the mean in a set. It considers all the values of a set, instead of just the extreme values, and calculates how close the values are spread around the mean. It is calculated as the average of the squared deviations of each value from the mean.

2. Standard Deviation

Standard deviation (SD) provides the measure of dispersion from the mean. It is the square root of the variance. Therefore, variance and standard deviation can be denoted as, $\mathrm{SD} = \sqrt{\mathrm{V}}$

How to Calculate Standard Deviation:
1. Calculate the mean.
2. Find the deviation of every value from the mean.
3. Square the deviations (referred to as squared deviations).
4. Calculate the average of the squared deviations.
5. Finally, take the square root of the average obtained in step 4. This is the standard deviation of the set.

   Standard Deviation = $\sqrt{\sum \frac{(\mathrm{x_i} - \mathrm{Mean})^2}{\mathrm{n}}}$,

   Variance = $\sum \frac{(\mathrm{x_i} - \mathrm{Mean})^2}{\mathrm{n}}$

Some Properties of Standard Deviation:

- Standard deviation cannot be negative. It will be zero if there is only one value in the set, or if all the values are the same (similar to the range).
- Standard deviation remains unchanged due to the addition (or subtraction) of a constant to each value in a set. Also, it will remain the same if the sign of each value is reversed (in other words, each value is multiplied by –1).
- If a constant is multiplied by each member of a set, then the new standard deviation will be obtained by multiplying the absolute value of the constant to the initial standard deviation.
- If the absolute values of the deviations of each observation from the mean are equal (e.g., *k*), then the standard deviation of the observations is equal to *k*.

A.3.2.7 Skewness and Kurtosis

When a set of data is plotted on a graph, it is said to be the distribution of the data. For example, a data set with a size of 30 (i.e., the data set contains 30 elements) is plotted on a straight line by marking the position of each element on the straight line. The mean of the data set is determined and marked on the line. If the elements are evenly placed on both sides of the mean, the data is symmetrical. Sometimes, it may be that a concentration of most of the elements is either to the left or the right of the mean. In such a case, the data is said to be skewed.

Distribution of a set of data is typically expressed by its shape, central tendency (average), or spread (standard deviation). It describes the nature of the data location: whether the data is concentrated near the mean or on either the left or right side of the mean (skewed).

The shape of the distribution is determined by attributes such as symmetry, skewness, and kurtosis.

1. Skewness

Skewness is a measure of asymmetry in a data distribution. If a data distribution shows a tendency to cluster around the higher values, then it is negatively skewed and, if it shows a tendency to cluster around the lower values, then it is positively skewed. Skewness is defined as the ratio of third central moment (μ_3) to standard deviation cubed.

$$Skewness = \frac{\mu_3}{(Standard\ deviation)^3},\ where\ \mu_3 = \frac{1}{n}\sum(x_i - Mean)^3$$

There are other measures of skewness including simpler calculations suggested by Karl Pearson.

Pearson's first skewness coefficient = $\frac{mean - mode}{standard\ deviation}$

Pearson's second skewness coefficient = $\frac{3(mean - median)}{standard\ deviation}$

For a symmetrical distribution, skewness is equal to zero, because mean, median, and mode are the same for a symmetrical distribution. If the calculated value of skewness is very close to 0, it can be inferred that the data distribution is approximately symmetric.

For highly skewed data, the median would be either left or right of the mean, and there would be some distance between them. The larger the distance between the mean and median, the more skewed the data will be, as can be verified by Pearson's formula.

2. Kurtosis

Kurtosis is a measure of the peaked-ness or flat-ness of the data distribution. Kurtosis (β_2) is defined as the ratio of fourth central moment (μ_4) to variance.

$\beta_2 = \frac{\mu_4}{Variance}$, where $\mu_4 = \frac{1}{n}\sum(x_i - Mean)^4$.

If a large portion of the data is concentrated near the mean and very few elements are away from the mean, then the variance will be very small. In this case, kurtosis will be very high and the distribution graph will have a high peak. On the other hand, if most of the elements of the data set are far from the mean (i.e., concentration near the mean is less), then the distribution graph will be flat.

A.3.2.8 Concept Testers

1. Find the standard deviation of the set: {0, 1, –1, 3, 4, 4, 2, –3, –3, 3}.
   Answer: 2.5

   Justification:
   a. Calculate the mean that is equal to (0 + 1 – 1 + 3 + 4 + 4 + 2 – 3 – 3 + 3)/10 = 10/10 = 1.
   b. Calculate the deviations (x - Mean) of every value from the mean (shown in the table below).
   c. Calculate the square of the deviations $(x - Mean)^2$ (shown in the table below).
   d. Calculate the average of the squared deviations = 64/10 = 6.4.
   e. Calculate the square root of 6.4, which is the standard deviation.

f.

X	Mean	(x - Mean)	(x -Mean)^2
0	1	–1	1
1		0	0
–1		–2	4
3		2	4
4		3	9
4		3	9
2		1	1
–3		–4	16
–3		–4	16
3		2	4
Total = 10	**10/10 = 1**		**Total = 64**

SD	**2.5**	Standard Deviation (SD) = $\sqrt{\frac{64}{10}} = \sqrt{6.4} = 2.5$

2. Find the mode of the set: {2, 1, 2, 3, 3, 3, 4, 1, 2, 4, 5, 8, 2, 2, 1, 2, 1, 8, 1, 1}.
   Answer: 1 and 2

   Justification:
   a. Find out the distinct values of the set; they are 1, 2, 3, 4, 5, 8.
   b. Determine the frequency of each value.

X	Frequency
1	6
2	6
3	3
4	2
5	1
8	2

   Since both 1 and 2 have the highest frequency (6), the set has two modes, 1 and 2.

3. Calculate the difference between the mean and median of the set: {3, 1, –3, –1, 5, -5}.
   Answer: 0

Justification:
If the values of the set are arranged in ascending order, they appear as –5, –3, –1, 1, 3, 5. Note that the values are evenly spaced since the common difference between any two consecutive values is 2. Therefore, the mean is equal to the median for this set.

4. Find the range of the set: {23, –22, 15, 0, 2, 30, –21, 3, 8, 11, –2}
Answer: 52

Justification:
Range = Highest value – Lowest value = 30 – (– 22) = 30 + 22 = 52

5. If a set of 7 values totals 35, and a new value (5) is added to the set, what will be the percentage change to the new mean?
Answer: 0

Justification:
Because the initial mean $= \frac{35}{7} = 5 =$ the value added, there will be no change to the new mean; in other words, the new mean will remain the same. Thus, the percentage change is zero.

Recall from section A. 2.1 that the mean can remain unchanged in two situations:
- If the initial mean is equal to the value of the member(s) added or deleted
- If all the members of the set have equal values as that of the added or deleted member(s)

6. Find the median of the set {15, 20, 1, 7, –5, 32, 16, 8}.
Answer: 11.5

Justification:
Arrange the values of the set in ascending order: –5, 1, 7, 8, 15, 16, 20, 32
Since the number of values is 8, which is an even number, the median will be the average of the two middle values

Thus, the median $= average\ of \left(\frac{8}{2}\right)^{th}\ and\ \left(\frac{8}{2}+1\right)^{th}\ values$

$= average\ of\ 4^{th} and\ 5^{th} values$

$$= average\ of\ 8\ and\ 15 = \frac{8+15}{2} = \frac{23}{2} = 11.5$$

7. If the standard deviation of a given set of 10 values is 3.5, what will the new standard deviation be if all the values of the set are multiplied by –2?
Answer: 7

Justification:
The new standard deviation will be 3.5 x absolute value of (– 2) = 3.5 x 2 = 7.

Recall from section A.2.8 that if we multiply a constant value to each member of a set, then the new standard deviation will be obtained by multiplying the absolute value of the constant to the initial standard deviation.

8. If 25% of the values of a given set are greater than or equal to the mean, then which of the following is NOT true? [Assume that the values of the data set are all distinct.]

   A. Mean = Median
   B. Mean > Median
   C. Mean < Median

   Answer: C. Mean < Median

   Justification:
   Since we know that 50% of the data values are greater than or equal to the median, and it is given that only 25% of the data values are greater than or equal to mean, the mean must be greater than or equal to the median assuming that the values of the data set are all distinct.

9. Which of the following remains the same due to the addition of a constant to each value of a set?

   A. Mean
   B. Median
   C. Mode
   D. Standard deviation
   E. Range

   Answer: Both D and E

   Justification:
   Refer to the properties of standard deviation and range.

10. What is the sum of the deviations of each value of a set from its mean?
    Answer: 0

    Justification:
    The sum of the deviations of each value of a set from its mean is always 0. Let the set be {x_1, x_2, x_3, …, x_n}, where n is a positive integer. Then, the sum of the deviations of each value from the mean = Σ (x_i – Mean) = Σ x_i – (n x Mean) = n x Mean – n x Mean = 0.

    Recall the definition of mean. Mean $= \frac{sum\ of\ values}{count} = \frac{\sum x_i}{n}$
    Multiplying both sides of the above equation by n, is $\sum x_i = n \times Mean$

11. Find the mean of the first 5 prime numbers.
    Answer: 5.6

    <u>Justification:</u>
    The first 5 prime numbers are 2, 3, 5, 7, and 11. The mean $= \frac{2+3+5+7+11}{5} = \frac{28}{5} = 5.6$

12. Which of the following can be negative?

    A. Mean
    B. Median
    C. Mode
    D. Standard deviation
    E. Range

    Answer: A, B, and C

    <u>Justification:</u>
    Standard deviation and range cannot be negative. Mean, median, and mode can take any value (positive, negative, or zero) depending on the data set.

13. A set can have more than one _____________. Fill in the blank from the following options.

    A. Mean
    B. Median
    C. Mode
    D. Standard deviation
    E. Range

    Answer: C

    <u>Justification:</u>

    The mode of a set is the value with the highest frequency of occurrence. A set may have more than one mode if several values have the same frequency of occurrence.

A.3.3 Probability

Probability can be defined as the possibility or likelihood of some event occurring. It is generally expressed in terms of a percentage, decimal, or fraction. The range of probability is from 0 to 1 (inclusive).

If an event is certain to occur (i.e., it will definitely happen), then the probability is said to be 1 (or 100%). For example, if a die having all sides red is rolled, the probability that the top side will be red is 100%.

On the other hand, if an event is certainly not possible to occur under the present conditions, the probability of that event is said to be 0 (or 0%). For example, the probability that a person's date of birth is February 29, 2014, is 0%. The date February 29, 2014, does not exist because 2014 was not a leap year.

A.3.3.1 Some Terminology in the Context of Probability

This section defines some terminology in the context of probability with some examples to create a foundation for this discussion.

1. Trial

Any particular performance of a random experiment is called a trial.

Examples of Trial:

- Tossing a fair coin
- Rolling a die
- Drawing a ball from a basket consisting of four white balls and six blue balls

Often the words *trial* and *experiment* are used synonymously, but when it is required to differentiate between these two terms, consider the experiment to be a larger entity formed by a combination of two or more trials.

In the experiment of tossing a coin four times, each tossing of the coin is considered a trial that is repeated four times under identical conditions.

2. Outcome

The result obtained from a single trial of an experiment is called an outcome.

Example of Outcome:

- The possible outcomes are heads or tails for the trial of tossing a fair coin.

3. Sample Space

The set of all possible outcomes of a random experiment is said to be the sample space for that experiment. It is denoted as *S*.

Examples of Sample Space:

- For the experiment of rolling a die, the sample space is *S* = {1, 2, 3, 4, 5, 6} (i.e., all of the six sides numbered 1 to 6).
- For the experiment of tossing a fair coin two times under identical conditions, the sample space is S = {(H, H), (H, T), (T, H), (T, T)}. Here, *H* and *T* represent heads and tails, respectively.

  Let's illustrate the second example. When a fair coin is tossed twice under identical conditions, the outcome will be in the form of heads or tails, heads or tails (H or T, H or T). Since the experiment under consideration is a combination of two trials (tossing a fair coin) and each trial results in two outcomes, the total number of possible outcomes for the experiment is = 2 x 2 = 2^2 = 4.

  The concept can be generalized to derive a useful formula for solving problems. If a fair coin is tossed *n*-times under identical conditions, the number of possible outcomes = $n(S) = 2^n$, where n(S) = the number of elements in the set S.

4. Event Space

An event space is a subset of the sample space. It is the set corresponding to the event under consideration. It is denoted as *E*.

Example of Event Space:

- In the previous problem of tossing a fair coin twice under identical conditions, an event can be defined as getting a single 'heads.'" Therefore, there should be only one heads and only one tails. The heads can occur in the first toss or in the second toss. Therefore the event space is *E* = {(H, T), (T, H)}.

Probability can now be defined more mathematically.

A.3.3.2 Probability of a Single Event

The probability of an event *E* can be defined as P(*E*) = *n*(*E*) / *n*(S). In other words, the probability of an event is defined as the number of favorable outcomes (when *E* occurs) divided by the total number of possible outcomes.

In the context of the previous problem, the probability of getting a single "heads" when tossing a fair coin twice under identical conditions needs to be found. Let *E* be the event "getting a single 'heads.'" Thus, P(*E*) = *n*(*E*) / *n*(S) = 2/4 = 1/2.

A.3.3.3 Important Definitions Regarding Two Events

It is easy to determine the probability of a single event, but it becomes complicated when dealing with the probability of two events. This section discusses some concepts involving two events to deal with this type of situation.

1. **Calculating the Probability of One Event or Another—The General Addition Formula**

   Consider two events: A and B. The probability of either of the events occurring (A or B) is given by the general formula P(A or B) = P(AUB) = P(A) + P(B) – P(A∩B).

   Note that to find the probability of A or B, it is necessary to find the probability of A and B (i.e., P(A∩B)).

   This formula is similar to the one used in Set Theory. In Set Theory the formula n(AUB) = n(A) + n(B) – n(A∩B) is used. Thus, most of the problems can be solved using the concepts from Set Theory.

Example of Calculating the Probability of One Event or Another Occurring:

- In a sample group, 40% of the subjects like to play basketball, while 70% like to play baseball. If 15% of the subjects like both of the games, to find the probability that a single person likes either of the games, one would take the following steps:

  Define the events A and B as "an individual likes to play basketball" and "an individual likes to play baseball," respectively. In this scenario, P(A) = 0.4, P(B) = 0.7, P(A∩B) = 0.15.

  Therefore, the probability of an individual liking either game = P(AUB) = P(A) + P(B) – P(A∩B) = 0.4 + 0.7 – 0.15 = 0.95

2. Mutually Exclusive Events

Two events are said to be mutually exclusive if they can never occur simultaneously. Mathematically, two events, A and B, are mutually exclusive if $A \cap B = \{\ \}$.

Therefore, if two events A and B are mutually exclusive, the following results:

- $P(A \text{ and } B) = P(A \cap B) = 0$
- $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$

Examples of Mutually Exclusive Events

- A person travelling to Phoenix and Washington at the same time
- A person scoring the highest and lowest marks on the class test

3. Complementary Events

An event A is said to be a complement of event B if one and only one of them must occur, but they cannot occur simultaneously. All complementary events are mutually exclusive. When these two criteria apply, A and B are said to be complementary events.

Examples of Complementary Events

- Winning the annual basketball match or losing the match
- Getting an odd face value on rolling a die or getting an even face value

Some Properties of Complementary Events:

- All complementary events are mutually exclusive, but the converse is not true (i.e., two mutually exclusive events need not necessarily be complementary to each other).
- The corresponding sample space can be obtained by combining the event spaces of the complementary events. In other words, adding the probabilities of two complementary events results in 1. Therefore, if A and B are two complementary events, $P(A) + P(B) = 1$ or, $P(B) = 1 - P(A)$.

4. Independent Events

Two events A and B are said to be independent if the occurrence of A is not dependent on the occurrence of B and vice versa.

If A and B are two independent events, then $P(A \cap B) = P(A)P(B)$ (i.e., the probability of the occurrence of both the events is obtained by multiplying the individual probabilities.)

Examples of Independent Events:

- Rolls of a die
- A job candidate getting selected by two different companies (assuming the job interviews occurred at different times)
- Drawing balls from a box containing identical balls with replacement (Since the ball is placed back in the box after the first draw, the number of balls remain the same for the second draw. Thus, the probability of the second draw is independent of the first draw.)

5. Dependent Events

Two events A and B are said to be dependent if the occurrence of A depends on the occurrence of B and vice versa.

Examples of Dependent Events:

- Drawing balls from a box containing identical balls without replacement (Since, the ball is not placed back in the box after the first draw, the number of balls is reduced before the second draw. Thus, the probability of the second draw is dependent on the first draw.)
- Picking cards out of a deck without replacement

6. Conditional Probability

The conditional probability of an event A, given that another event B has already occurred, is defined by

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

Examples of Conditional Probability:

- Cards are drawn from a pack of 52 cards, one by one without replacement. What is the probability that the third draw will be a red card given that both cards were black in the first and second draw?

  Since cards are drawn without replacement, there will be 50 cards after the second draw. Both the cards were black in the first and second draw, so there will be exactly 26 red cards before the third draw. Therefore, the probability will be $\frac{26}{50}$.

A.3.3.4 Concept Testers

1. Find the probability of getting at least one "tails" on tossing a fair coin twice.
   Answer: 0.75

   Justification:
   The sample space S = {(H, H), (H, T), (T, H), (T, T)}
   Let *E* be the event "getting at least one 'tails'"
   Therefore, the event space *E* = {(H, T), (T, H), (T, T)}
   Therefore, the required probability P(*E*) = *n*(*E*) / *n*(S) = 3/4 = 0.75

14. If it is given that P(A) = 0.2, P(B) = 0.7, P(AUB) = 0.65, find P(A and B).
    Answer: 0.25

    Justification:
    Apply the general addition formula: P(AUB) = P(A) + P(B) – P(A∩B)
    ⇒ P(A∩B) = P(A) + P(B) – P(AUB)
    ⇒ P(A and B) = 0.2 + 0.7 – 0.65 [since P(A and B) = P(A∩B)]
    ⇒ P(A and B) = 0.25

15. If A and B are mutually exclusive events, then find P(A or B), given that P(A) = 0.4 and P(B) = 0.35
    Answer: 0.75

    Justification:
    Since A and B are mutually exclusive events, P(A or B) = P(AUB) = P(A) + P(B) = 0.4 + 0.35 = 0.75

16. A jar contains 12 identically shaped and sized marbles, 4 of which are green and the rest are blue. If marbles are picked randomly with replacement, find the probability of the second draw being a green marble.
    Answer: 1/3

    Justification:
    Since it is given that the marbles are pulled from the jar randomly with replacement, the event *E* "second draw is a green marble" is independent of the first draw, whether it is green or blue. Note that after the first draw, the marble is replaced in the jar, and then the second marble is drawn. Therefore, the probability is equivalent to the probability of drawing a green marble from a jar containing 12 identical marbles, 4 of which are green and the rest are blue.

    Therefore, P(*E*) = 4/12 = 1/3

17. When a card is drawn from a standard deck of playing cards, what is the probability that the drawn card is either a black card or an ace?

Answer: $\frac{7}{13}$

Justification:
Let A be the event of "getting a black card" and B be the event of "getting an ace."

Since there are 26 black cards and 4 aces in a standard deck of playing cards, $P(A) = \frac{26}{52}, P(B) = \frac{4}{52}$

Since there are two black cards that are aces, $P(A \cap B) = \frac{2}{52}$
By the addition theorem on probability,
P(AUB) = P(A) + P(B) – P(A∩B) $= \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{26+4-2}{52} = \frac{28}{52} = \frac{7}{13}$

18. If A and B are two independent events such that P(A) = 0.25 and P(B) = 0.6, find P(A or B).
Answer: 0.7

Justification:
Since A and B are two independent events, P(A∩B) = P(A) P(B) = 0.25 x 0.6 = 0.15
Therefore, P(A or B) = P(AUB) = P(A) + P(B) – P(A∩B) = 0.25 + 0.6 – 0.15 = 0.7

A.3.4 Random Variables and Probability Distributions

A.3.4.1 Random Variables

In statistics, a random variable is allowed to choose any value randomly within its permissible range. For example, a statistician is interested in studying the birth rate of a city and collects data on the "number of babies born per hour" for the past three months in that city. Let X denote the random variable "number of babies born per hour in the chosen city." Therefore, X can take any non-negative integer value (i.e., possible values of X are 0, 1, 2, 3, 4, etc.).

The lowest possible value that X can take is zero because it is possible to have no baby born in a specific time period of 1 hour. The maximum possible value that X can take is any positive integer or whole number.

Random variables are broadly classified as "discrete" and "continuous." If the set of possible values of a random variable is a countable set, then the random variable is said to be a discrete random variable. A countable set is a set that can be counted. A countable set may be finite or infinite. Examples of countable sets include the following:

- any finite set of values
- the set of all whole numbers
- the set of all integers
- the set of all rational numbers
- any subset of whole numbers
- any subset of integers

If the set of possible values of a random variable is an interval or a collection of non-overlapping intervals, then the random variable is said to be a continuous random variable. An interval is defined as a set of all values ranging from the lowest number to the highest number. For example, an interval (2, 10) denotes a set of all values between 2 and 10 (which includes whole numbers as well as fractions). A few more examples of discrete and continuous random variables are given in Table A.3-1.

Discrete Random Variables	Continuous Random Variables
The random variable takes distinct or separated values. In other words, a discrete random variable takes values in a finite or countable set. The possible values of a discrete random variable are normally obtained by count. *Examples:* • X = number of "heads" while tossing a fair coin 5 times successively. • X = number of luggage lost per month for an airline company. • X = sum of upward face values of a fair die after rolling it 4 times successively.	The random variable takes any value in an interval. In other words, a continuous random variable takes values in an uncountable set. The possible values of a continuous random variable are normally obtained by a physical instrument. *Examples:* • Y = the exact weight of a newborn baby randomly selected from all the hospitals in a city. • Y = the exact time taken by the participants of a 100-meter race. • Y = the waiting time for a customer standing in a queue to receive a service.

Table A.3-1: Discrete and Continuous Random Variables

A.3.4.2 Probability Distribution of a Random Variable

Consider the following example showing how to construct the probability distribution of a discrete random variable. Let X be the random variable defined as X = sum of upward face values of a fair die after rolling it 3 times successively. Since a fair die has 6 faces, each marked with values 1 to 6, the minimum sum will be 1+1+1 = 3 and the maximum sum will be 6+6+6 = 18. Therefore, the random variable X can take all the integer values from 3 to 18, both inclusive. This is a discrete random variable because X takes values in a finite set {3, 4, 5, 6… 17, 18}.

A probability distribution is a graph of the values of X and the corresponding probabilities. To create a distribution, determine the probabilities for each value of X. Let's find P(X = 5).

The total number of outcomes = 6 x 6 x 6 = 216 (this is true because for each roll there are 6 possible values, and we need to consider ordered arrangement).

Now consider how to obtain the sum of 5 from 3 successive rolls of a fair die. This is illustrated in Table A.3-2.

1st roll	2nd roll	3rd roll	Total
1	1	3	5
1	3	1	5
3	1	1	5
1	2	2	5
2	1	2	5
2	2	1	5

Table A.3-2: Sample Results for Probability Distribution

It is clear that there are 6 possible ways to arrive at X = 5. Thus, P(X = 5) = $\frac{6}{216} = \frac{1}{36}$. Similarly, other probabilities can be determined, and a graph of values of X and corresponding probabilities can be drawn.

A.3.4.3 Discrete Probability Distributions

Some of the more common discrete probability distributions and their applications are provided in this section. These are as follows:

- Binomial
- Poisson
- Hypergeometric

A.3.4.3.1 Binomial Distribution: Bin (n, p)

A binomial distribution is a discrete probability distribution of the number of successes in a sequence of success/failure or yes/no experiments (e.g. tossing of a fair coin n times under identical conditions). A Bernoulli trial is an experiment that has only two possible outcomes—success (s) and failure (f).

A sequence of *n* Bernoulli trials has the following properties:

- The trials are independent of one another: the outcome of one trial does not depend on the outcome of any other trial.
- The trials are identical. P(Success) = constant for all trials. Let us denote this by “p.”

Such a sequence is called a sequence of "n" independent and identical Bernoulli (p) trials. This is known as Binomial (n, p) distribution.

Conditions for the use of Binomial Distribution:

- Events are independent.
- Events are mutually exclusive with two outcomes (Success/Failure, Good/Defective, Yes/No, etc.).
- Probability (p) of each outcome should remain constant during trials.
- Number of trials is finite (n) for a particular experiment and n can take any non-negative integer value.

A.3.4.3.2 Poisson Distribution – Poi (λ)

A Poisson distribution is the number of events that occur in a specified time interval with the following assumptions:

- Events occur singly (one after another) at a constant rate, λ per unit of time.
- The number of events that occur in non-overlapping time intervals is independent of one another.
- The events are described as occurring as a "Poisson process with rate λ."

Poisson distribution describes the number of events that occur in a specified interval of time.

If a sequence of binomial (n, p) distributions are such that the mean np is held constant as λ with the two limiting conditions $n \to \infty$ and $p \to 0$ being satisfied simultaneously (typically $n \geq 100$ and $p \leq 0.05$), then the limit leads to the Poisson (λ) distribution.

A.3.4.3.3 Hypergeometric Distribution

Hypergeometric distribution is similar to binomial distribution but with finite population and without replacement.

Let the objects be selected at random, one after another without replacement from a finite population of size N consisting of k successes and N - k failures. Here, the trials are not independent since the selection is made without replacement.

The mean is given by nk/N (note that in the Binomial (n, p) distribution, the mean is given by np).

The binomial p = k/N provides a good approximation to the hypergeometric distribution in many cases.

A.3.4.4 Continuous Probability Distributions

As abovementioned, when the set of possible values of a random variable is an interval or a collection of non-overlapping intervals, then the random variable is said to be a continuous random variable. Consider the following example showing the probability distribution of a continuous random variable.

Let X be a continuous random variable. Since X is continuous, it will take values in an interval rather than taking discrete values. The probability distribution of X is made of a continuous function called a "probability density function."

Probability Density Function (PDF)

The probability density function (PDF) is a continuous mathematical function, generally denoted as f(x) that models the probability density reflected in a histogram. PDF suggests the shape of the probability distribution.

The PDF uses integrals—a summation of all the areas under the curve. Figure A.3-1 shows a graphical representation of the probability distribution of X where the horizontal axis represents the values of X and the vertical axis represents the values of the PDF curve f(x).

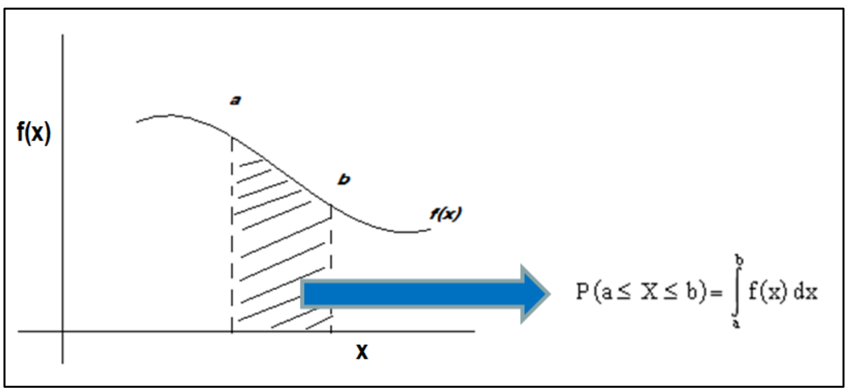


Figure A.3-1: Sample Continuous Probability Distribution

This PDF curve is used to determine the probability of the value X in a specified range. The mathematical formula used for this is the following:

$$P(a < X < b) = \int_a^b f(x)dx$$

This means that the probability that the random variable X has values ranging from *a* to *b* is given by the area under the PDF curve. The area is enclosed by $x = a$ and $x = b$ under the PDF curve as shown in the diagram above. Note the difference between the construction of a discrete and continuous random variable—in the case of a discrete random variable, $P(X = a)$ (i.e., the probability of X = a single value *a*);

however, in the case of a continuous random variable, as shown in the formula above $P(a < X < b)$ (i.e., the probability of X lies in the interval (a, b)).

Some of the more common continuous probability distributions and their applications are provided in this section. These are as follows:

- Normal
- Lognormal
- Exponential
- Weibull
- Chi-square
- Student's Distribution
- F Distribution

A.3.4.4.1 Normal Distribution: N(μ, σ)

The graph of a normal distribution looks like a bell curve that is symmetric about the mean μ as shown in Figure A.3-2.

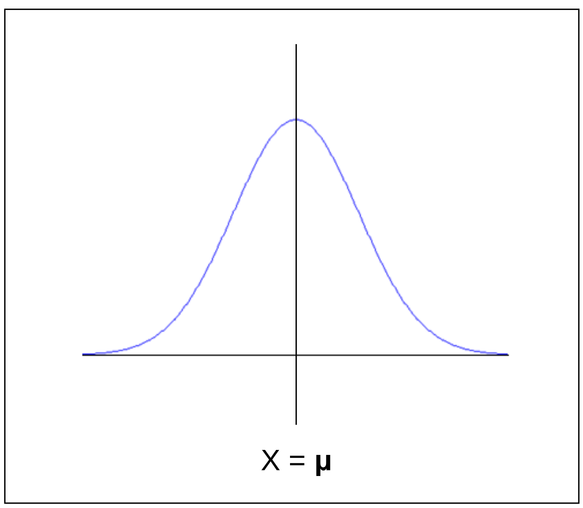


Figure A.3-2: Normal Distribution

Properties of a normal curve include the following:

- It is bell shaped (i.e., most of the data points are concentrated near the mean).
- It is symmetric about the mean.
- Mean = median = mode.
- The area under the normal curve = 1.
- It is unimodal (i.e., has one mode).

Many naturally occurring random processes tend to have a distribution that is approximately normal. The SAT test scores of college-bound students and body temperatures of a healthy adult both follow a normal distribution. As per the Central Limit Theorem, the means of sample observations (i.e., sample means) are approximately normal if the sample sizes are large enough, whether the population is normally distributed or not. The assumption of normality of the population is very important for conducting additional statistical analyses (e.g., testing of hypothesis and correlation and regression analysis). Even if the distribution of the population is unknown or is not normally distributed, we can use a large sample to apply the Central Limit Theorem and allow the sample means to follow normal distribution approximately.

Standard Normal Distribution N(0, 1)

A normal distribution with mean μ and standard deviation σ can be made standard normal distribution using a linear transformation $Z = (X - \mu)/\sigma$. If the random variable X represents the normal distribution with mean μ and standard deviation σ, then the random variable Z will represent a standard normal distribution. The standard normal distribution has a mean of 0 and a standard deviation of 1. This is known as standardization of a normal distribution. It is required for further statistical analysis because it makes the calculations easy.

A.3.4.4.2 Lognormal Distribution

A lognormal distribution is a right-skewed distribution with most of the data residing in the left tail. For this distribution, natural logarithms of the original data follow a normal distribution. Figure A.3-3 shows a sample lognormal distribution.

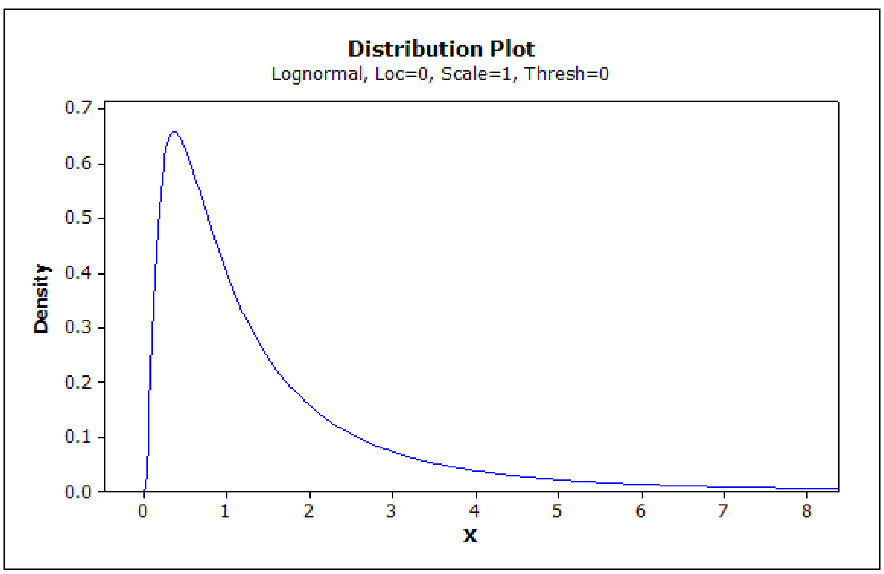


Figure A.3-3: Sample Lognormal Distribution

A.3.4.4.3 Exponential Distribution

Exponential distribution describes the time between events in a Poisson process (i.e., a process in which events occur continuously and independently at a constant average rate). The number of events in a Poisson process is modeled as a Poisson distribution, whereas the waiting time between the events is modeled as an exponential distribution. It has the key property of being memoryless. It is used to model lifetimes of certain elements, to determine the average time between failures or the average time between a number of occurrences, and to analyze reliability. Figure A.3-4 provides a sample exponential distribution.

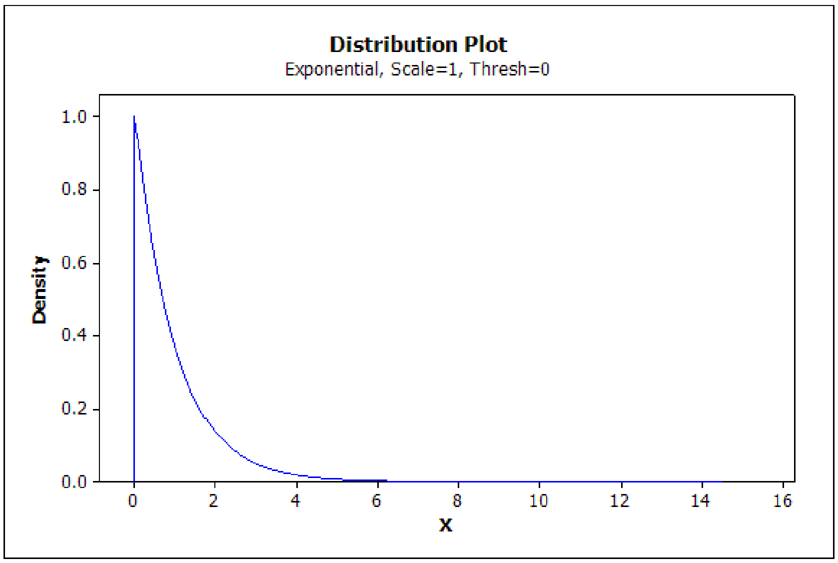


Figure A.3-4: Sample Exponential Distribution

A.3.4.4.4 Weibull Distribution

The Weibull distribution is used in analyzing reliability. It can also be used in applications that are similar to those that use a lognormal distribution (e.g., to measure such values as time to fail, time to repair, and material strength).

A.3.4.4.5 Chi-squared (χ^2) Distribution

A Chi-squared distribution is a right-skewed distribution: the curve has its long tail toward the higher values of the distribution. Some characteristics of this distribution include the following:

- Tests hypotheses regarding population variance against known or assumed value
- Assesses two types of comparison:
    - Chi square test for goodness of fit of an observed distribution to a theoretical one

  - Test of independence, which assesses whether paired observations on two variables, expressed in a contingency table, are independent of each other
- Has one parameter *k*, which is a positive integer that specifies the number of degrees of freedom. Degrees of freedom is defined as an unrestricted variable (i.e. free to vary) in a sampling distribution. This variable can take any positive integer value.

Figure A.3-5 provides a sample Chi-squared (χ^2) distribution.

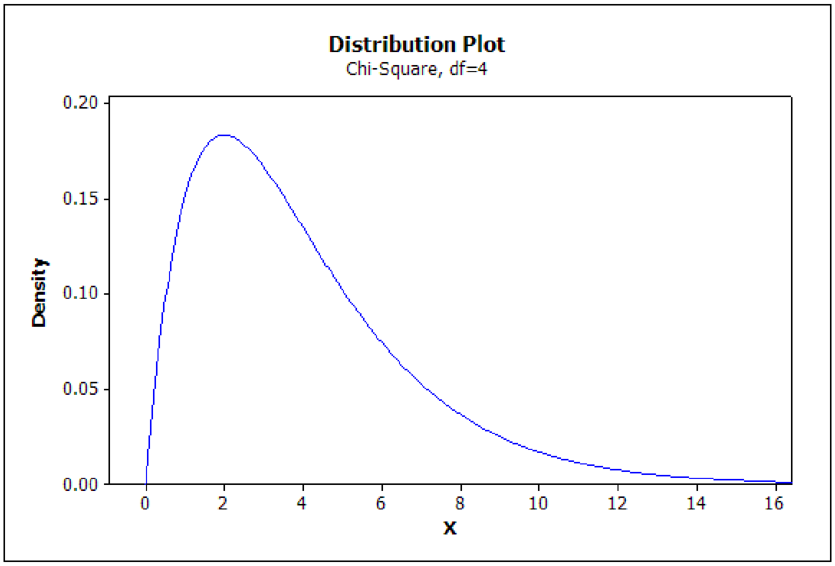


Figure A.3-5: Sample Chi-squared (χ^2)) Distribution

A.3.4.4.6 Student's t Distribution

Student's t distribution or simply "t distribution" is commonly used to determine the confidence interval of the population mean when population variance is unknown. Some characteristics of this distribution are as follows:

- It is used to test a hypothesis when the means of sample populations are compared
- The shape of the t distribution approaches the standard normal distribution as the sample size (n) increases
- Typically at $n \geq 30$, t and normal distribution will be the same
- t distribution has one parameter n, which is a positive integer that specifies the number of degrees of freedom

Degree of freedom is defined as an unrestricted variable in a distribution). It can be derived for a sampling distribution say *t*-distribution as described below. Degrees of freedom is the number of independent pieces of information required to estimate a population parameter. So when estimating population variance by using sample variance computed based on a sample size n there will be n – 1 degrees of freedom for sample variance. This is because there is one loss of degrees of freedom as explained below:

Random sample of size *n*: $x_1, x_2, \dots, x_n$

Sample variance = $\sum \frac{(x_i - \bar{x})^2}{n}$ where $\bar{x}$ is the sample mean calculated as $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$

$x_i - \bar{x}$ are the deviations of sample units from the sample mean for $i = 1,2, \dots, n$

If the sum of these *n* deviations are taken, the sum will be zero, i.e. $\sum(x_i - \bar{x}) = 0$

Hence it is sufficient to know (n – 1) of these deviations as the last one can be obtained using above formula. In other words there are exactly (n – 1) independent pieces of information which are required to estimate population variance. Since sampling distribution of sample mean is *t*-distribution which is derived based on sample variance, (n – 1) degrees of freedom is used for *t* statistic.

A.3.4.4.7 F Distribution

F distribution is used to test a hypothesis of equality of two population variances. Other characteristics of this distribution include the following:

- It is important in ANOVA, which is a technique used in design of experiments (DOE) for testing significant differences in variance within and between test runs
- F distribution has two parameters n_1 and n_2 called degrees of freedom
- The shape of the distribution curve is non-symmetrical and varies with the degrees of freedom

A.3.5 Sampling Distributions

Suppose from a given population, we are drawing all possible samples of size *n*. For each sample, we compute a sample statistic (e.g., mean, standard deviation, and proportion). The probability distribution of this statistic, considered a random variable, is called a sampling distribution of that statistic. These are considered a sampling distribution because they consist of sample statistics.

The sample statistic acts as a random variable because every time we draw a random sample of size *n* from the given population the statistic varies.

Let's us consider the sampling distribution of a sample mean for a random sample of size *n* drawn from a normal (μ, σ^2) population.

Let X be the random variable representing the population. The probability distribution of X has a normal distribution with population mean μ and variance σ^2. Let a random sample of size *n* be drawn from the population and sample mean is calculated.

The probability distribution of the sample mean $\overline{X}$, as a random variable has a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

If we take 100 samples of size 15 from a population and find the means of all the samples, the resulting distribution is an example of the sampling distribution of mean.

The mean of this distribution is the same as the mean of the population, but the standard deviation of this distribution is smaller than the standard deviation of the population. This relationship is given by $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, where $\sigma_{\bar{X}}$ is the standard deviation of the sampling distribution of mean, σ is the standard deviation of the given population and *n* is the size of the sample. This is known as Standard Error.

A.3.5.1 Standard Error

The standard deviation of the sampling distribution of means indicates the standard error (SE), which is the amount of error that can occur when a sample mean is used for estimating a population mean. Standard error is calculated as standard deviation divided by the square root of *n* where *n* is the sample size.

Standard Error = $\frac{standard\ deviation}{\sqrt{n}}$

As sample size increases, the standard error decreases. In other words, the larger the sample size, the smaller the standard error; thus, estimation becomes more accurate.

So far, we have assumed the population to be normally distributed and based on that assumption we have determined the sampling distribution of mean. But in more practical situations, the population may not be normally distributed. What should we do in those situations? This can be explained by the Central Limit Theorem.

A.3.5.2 Central Limit Theorem (CLT)

The CLT states that when plotting the mean values of the samples taken from any population (which does not need to be normal), the distribution of the sample means tends to a normal distribution, as the sample size increases.

Figure A.3-6 shows the sample size and the spread of a sampling distribution of mean:

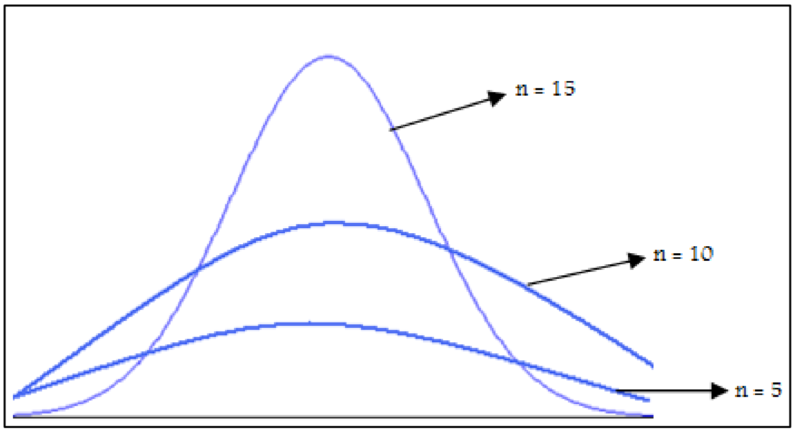


Figure A.3-6: Example of Distribution of Means for Growing Sample Size

The spread of the sampling distribution of mean decreases when the sample size increases. Therefore, CLT can be summarized as follows: Given any distribution with mean μ and variance σ^2, the sampling distribution of mean is approximately normally distributed with mean μ and variance $\frac{\sigma^2}{n}$, if the sample size *n* is sufficiently large.

What makes a sample large enough for the normality assumption to not be required? That depends on the nature of the data. If the data is highly skewed, then sample size should be as high as possible. But when the data is nearly symmetrical, then a sample size greater than or equal to 30 is acceptable. It is an industry practice to take a sample size of at least 30. The number may differ on a case-to-case basis and also depends on the skewness of data.

A.3.5.3 Applications of Sampling Distributions

Sampling distributions are used to find the confidence interval of population parameters such as mean and variance. Sampling distribution of sample statistics (known as test statistic) is also used in the testing of a hypothesis. Table A.3-3 shows applications of sampling distributions.

Sample Statistic	Sampling Distribution	Confidence Interval
Sample mean $\bar{x} = \frac{1}{n}\sum x_i$	Exactly normal with mean μ and variance $\frac{\sigma^2}{n}$ when sampling from a $N(\mu, \sigma^2)$ population.	[1] $\overline{x} - z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right)$ Sampling from a normal population with known population variance(σ^2)
	Approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ when sampling from a population with mean μ and variance σ^2 for a sufficiently large sample. The population need not be normal (applying the concept of central limit theorem).	[2] $\overline{x} - z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right)$ Sampling from any population with known population variance(σ^2). Sample size should be sufficiently large. [3] $\overline{x} - z_{\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + z_{\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}}\right)$ Sampling from any population when population variance is not known. Sample size should be sufficiently large.
	When a simple random sample of size *n* is drawn from a normal population with mean μ but unknown variance, then the expression $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ has a t distribution with $n-1$ degrees of freedom(t_{n-1}). The sampling distribution of t statistic is used to determine the confidence interval of population mean μ when population variance σ^2 is unknown.	[4] $\overline{x} - t_{\frac{\alpha}{2},n-1} \left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + t_{\frac{\alpha}{2},n-1} \left(\frac{s}{\sqrt{n}}\right)$ Sampling from normal population when population variance is not known. This formula is used when the sample size is small.
Sample variance $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$	The expression $\frac{(n-1)s^2}{\sigma^2}$ has a chi-squared distribution with $n-1$ degrees of freedom (χ^2_{n-1}) when sampling from a $N(\mu, \sigma^2)$ population.	$\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right),n-1}}}$ Sampling from a $N(\mu, \sigma^2)$ population.

Table A.3-3: Applications of Sampling Distributions

A.3.5.4 Derivation of Confidence Interval for a Population Parameter

The formula for the endpoints of the confidence interval (CI) for population mean μ is: $\left(\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right), \bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$ when population standard deviation σ is known. When σ is unknown, it can be estimated by sample standard deviation s.

α = probability that the population mean does not fall in the interval (α-risk)

$1-\alpha$ = probability that the population mean is in the interval (confidence level)

$z_{\frac{\alpha}{2}}$ = the value from the Z-table (standard normal table) with an area of α/2 to its right

This formula is derived by taking a random sample of size *n* from a population having $N(\mu, \sigma^2)$ distribution, in which the sampling distribution of sample mean $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n})$. Applying the standardization formula, we will have the random variable$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, which is a standard normal variant $N(0,1)$. In practical situations it may not be possible to have a normal population (i.e., the population from which samples are drawn is not normal). In such cases, the sampling distribution of sample mean $\bar{X}$ will be approximately $N(\mu, \frac{\sigma^2}{n})$ for a sufficiently large sample by virtue of the central limit theorem.

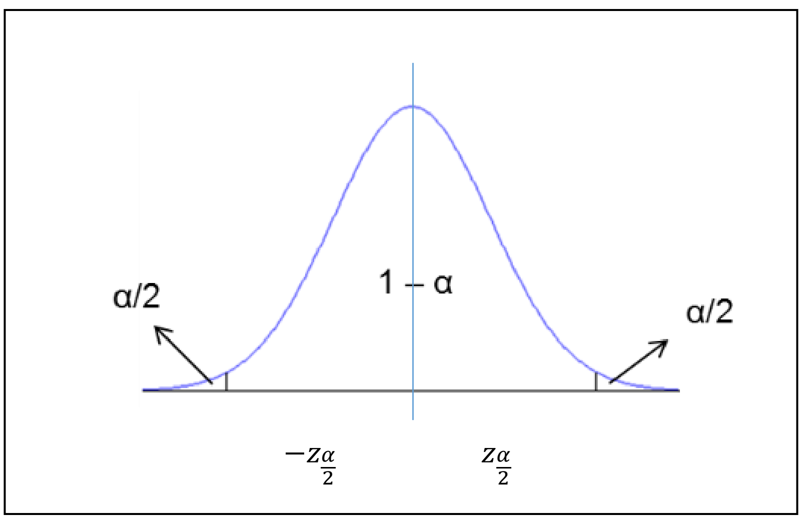


Figure A.3-7: Example Standard Normal Curve

Figure A.3-7 shows a standard normal curve having a mean of 0 and a standard deviation of 1. The notations are already defined. The probability that the random variable Z takes values between $z_{\frac{\alpha}{2}}$ and $-z_{\frac{\alpha}{2}}$

is the area under the standard normal curve enclosed by z = $z_{\frac{\alpha}{2}}$ and z = $-z_{\frac{\alpha}{2}}$. This is shown in the above diagram. Mathematically, this is written as $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$

Now we will work on the inequality, $-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$

Substitute, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ in the inequality to get $-z_{\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}$

Multiply $\frac{\sigma}{\sqrt{n}}$ on both sides of the inequality to get $-z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \bar{x} - \mu < z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

Multiply -1 on both sides of the inequality to get $z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) > \mu - \bar{x} > -z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$. Remember when we multiply by a negative number in an inequality, the inequality sign will reverse. The inequality can be written as

$$-z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu - \bar{x} < z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Add $\bar{x}$ on both sides of the inequality to get $\overline{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

Thus, the expression $P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$ can be written as

$$P\left(\overline{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$$

In this form, $1 - \alpha$ is called the confidence level and indicates how confident we are that the population's mean μ lies within the indicated confidence interval $\left(\overline{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right), \overline{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$. If the population standard deviation σ is not known, we can use the sample standard deviation s as an estimate. The confidence interval in that case will be $\left(\overline{x} - z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right), \overline{x} + z_{\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)\right)$. Please note that in both cases, we are using the standard normal variant. This is because either the population under study is normally distributed or we are using a sufficiently large sample to apply the central limit theorem.

A.3.6 Method of Least Squares

To understand the Least Squares Method, consider the following example.

Example of Method of Least Squares:

- A market research analyst is studying the quarterly sales figures with respect to expenses toward promotional campaigns of an organization. The researcher has collected data for the past four years for this study. A scatter plot indicates that there exists a positive linear relationship between these two variables: sales volume and expenses towards promotional campaigns. Now, the researcher wants to quantify this linear relationship between these two variables, (i.e., understand the equation of the straight line that can be described as "best fit" for these data). This line is known as the "Regression Line" in statistical terminology.

Through observation, a few lines through these points could be a good fit. Figure A.3-8 displays one solid line and three dotted lines, which the researcher tentatively thinks could be a good fit.

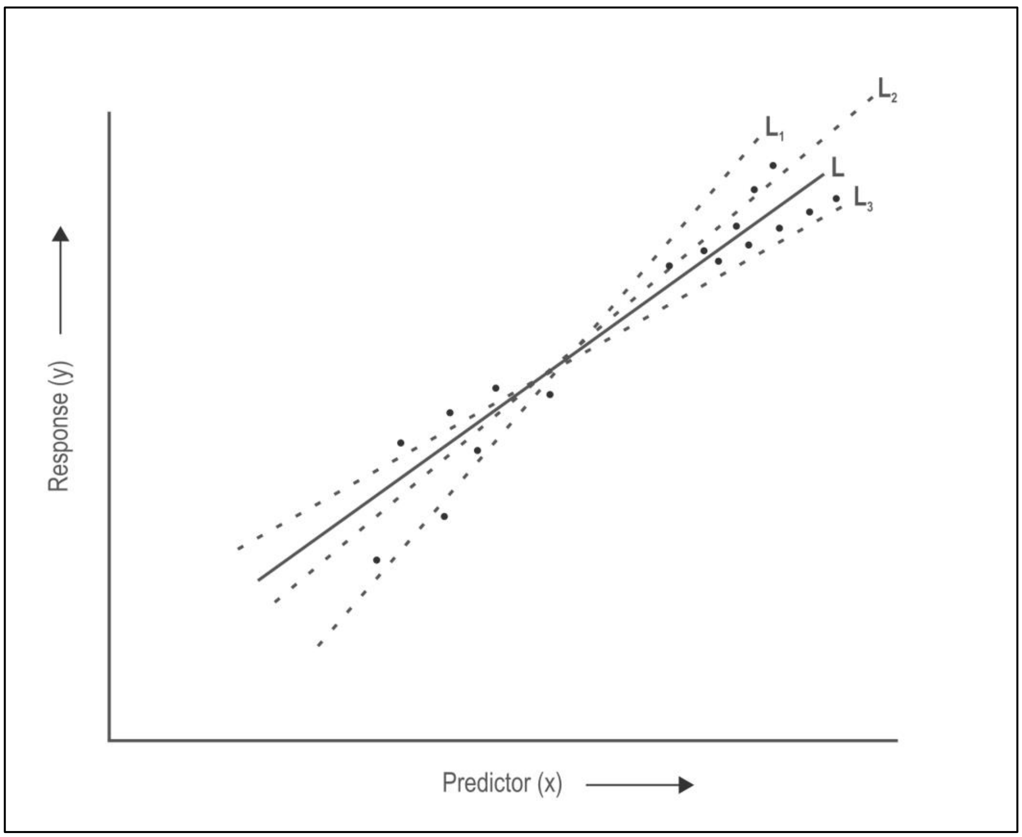


Figure A.3-8: Possible Regression Line "Good Fits" for the Example

In general, a good fit means that the line should be such that all the points will be as close as possible to the line. In other words, the distances of the points from the line are as minimal as possible. According to this logic, more than one line can be treated as a good fit. But the question is “which one is the best?” The answer to this question can be obtained by a statistical concept named “Method of Least Squares.” The theory states that among all these lines, the best fitted line is the one for which the sum of the squares of the distances of the points from the line is the minimum. Figure A.3-9 graphically displays the concept.

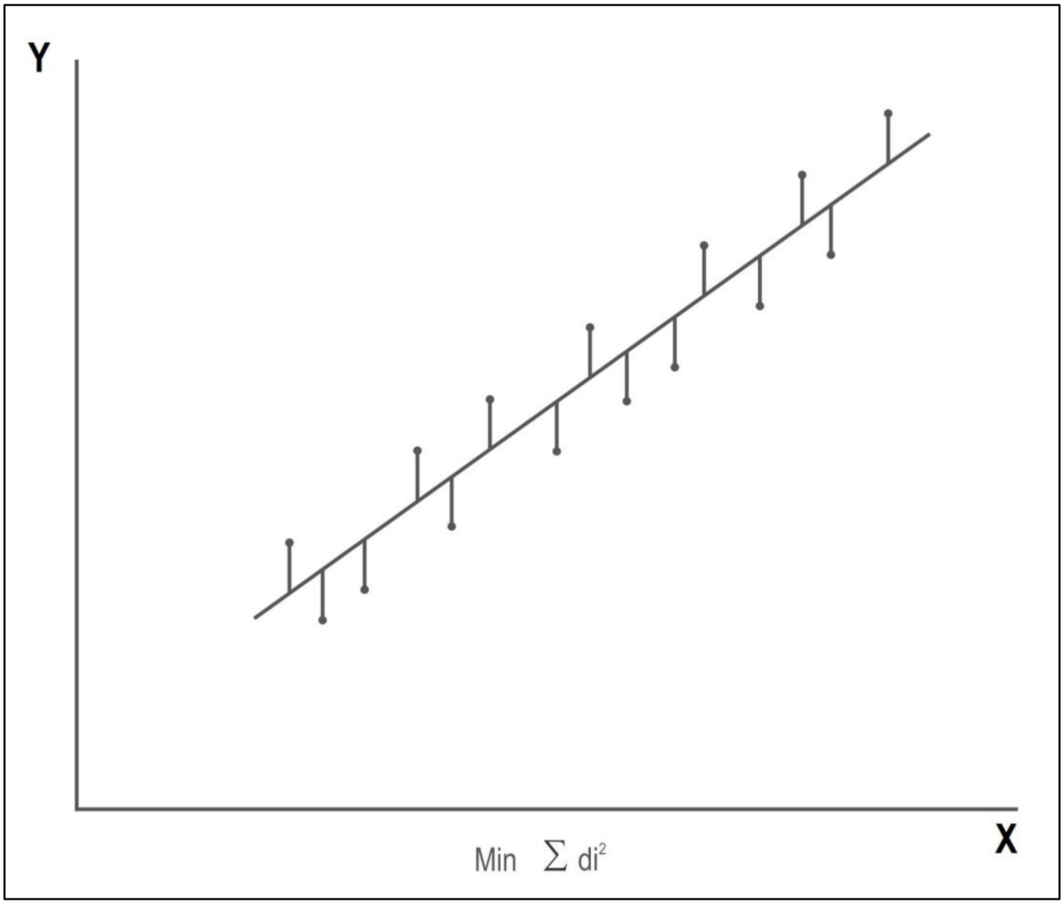


Figure A.3-9: Example of Method of Least Squares

The following are some notations and a graph for further clarification.

Denote the variable “sales volume” as Y. In statistical terms, this is called the “Response” variable.

Denote the variable “expenses towards promotional campaigns” as X. In statistical terms, this is called the “Predictor” variable.

To find a simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$, where β_0 is the y-intercept, β_1 is the slope of the line and ϵ is the random error term that is assumed to be normally distributed with zero mean and variance σ^2. For the time being ϵ ,β_0 and β_1 can be ignored.

In this example, we are examining promotional expenses as a predictor that drives sales. Therefore, promotional expenses are a factor that influences sales. There may be more factors that influence sales, but for simplicity we are examining only one predictor.

The data is represented as (x_i, y_i); where the subscript *i* denotes a serial number from 1 to 16. How have we arrived at the number 16? Recall that the data (quarterly sales vs. promotional expenses) was collected for the last four years. Each financial year comprised four quarters. Thus, there will be a total of (4 x 4) = 16 quarters in 4 years. If we name the quarters 1, 2 … 16 sequentially then x_i denotes promotional expenses and y_i denotes sales volume for the i^{th} quarter.

Figure A.3-10 shows the data points, (x_i, y_i). Also, note that $\hat{y}_i$ represents the predicted response or fitted value from the regression line corresponding to the predictor x_i.

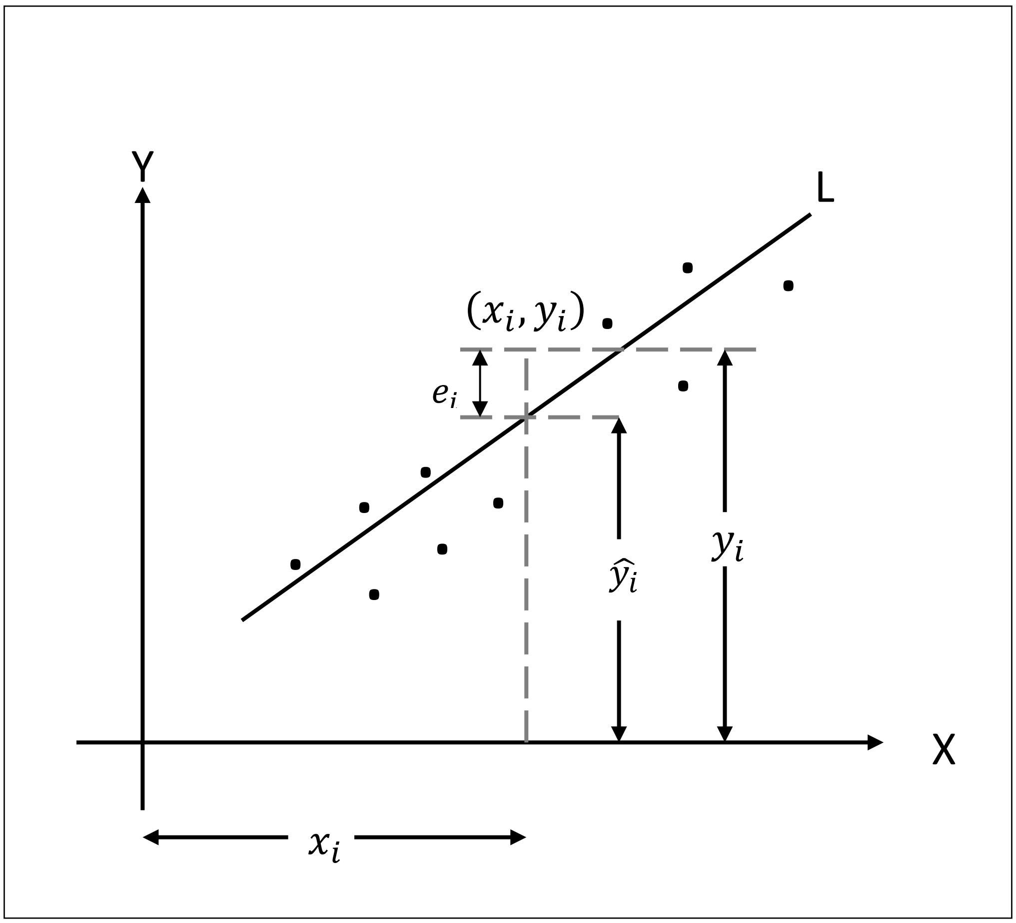


Figure A.3-10: Plotted Data Points for Sales Volume and Expenses

The line L represents the fitted line $\hat{y} = \beta_0 + \beta_1 x$.

The notation e_i represents the distance of the point (x_i, y_i) from the fitted line. It is also known as the residual or error, defined as $e_i = y_i - \hat{y_i}$ (i.e., for the fitted line $\hat{y} = \beta_0 + \beta_1 x$, response – fitted response = residual from the fit).

The least squares method is the technique of minimizing the sum of the squared errors (SSE) $\sum_1^n e_i^2$ = $\sum_1^n (y_i - \hat{y_i})^2$.

For a set of paired data (x_i, y_i); *i* = 1, 2… *n* the least squares estimates of the regression coefficient are the values "β_0" and "β_1," for which the sum $\sum_1^n e_i^2$ = $\sum_1^n (y_i - \hat{y_i})^2$ is minimum.

We need to calculate the residuals e_i for these 4 fitted lines. The next step is to calculate the sum of squared residuals (SSE). The best-fitted line will be the one for which the SSE is the minimum. Market researchers should not be concerned about the complicated mathematical formulas and calculations because any standard statistical package will perform the calculations: the market researcher's role is to do the analysis and interpret the result. These statistical packages have built-in algorithms that generate the best-fitted line within minutes.

A.3.7 Analysis With Statistical Packages

To demonstrate statistical analysis using statistical packages, the following scenarios are explored in this section using examples:

- One-way ANOVA
- Two-Way ANOVA
- Multiple Linear Regression
- Factor Analysis

Example of Statistical Packages:

- Statstical analysis can be performed using statistical packages such as Minitab, SPSS, and SAS.

Note: Statistical calculations in this section are performed and illustrated using Minitab version 17.

A.3.7.1 One-way ANOVA

Example One-way ANOVA (Single Factor ANOVA)

- A manufacturing company has three plants: X, Y, and Z. A random sample of scrap generated in pounds was collected for 6 days in the following manner. Do they differ in producing waste?

	Waste (in lbs)		
Observation No.	Plant X	Plant Y	Plant Z
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67

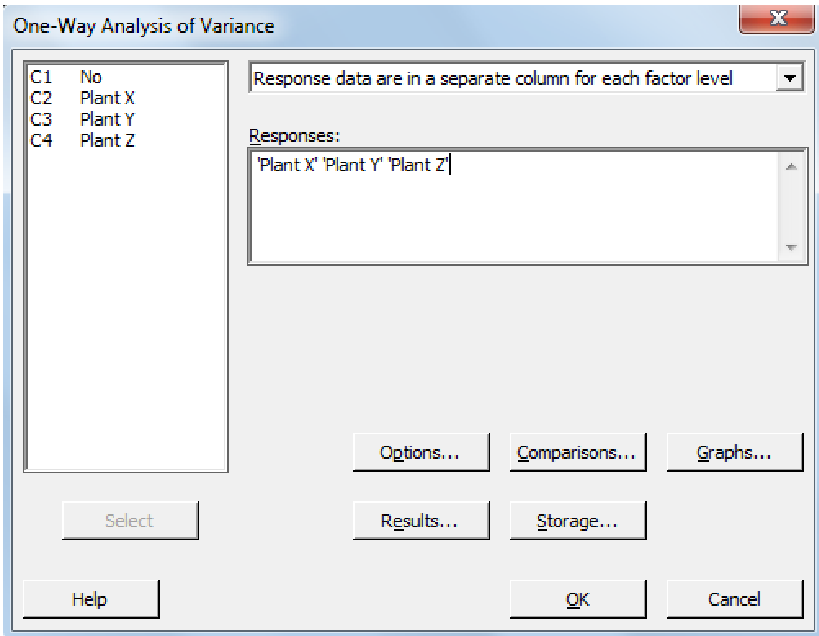
Assuming equal variances across plants, test whether or not the waste produced is the same at a 5% level of significance.
Therefore, we are testing the null hypothesis H_o: All means are equal against H_1: At least two of the means are different.

Copy the data in the Minitab worksheet the way it is given.

Select Stat → ANOVA → One-Way

When the dialog box appears, select "Response data are in a separate column for each factor level" from the drop down box.

Select the responses as Plant X, Plant Y, and Plant Z as shown in the below dialogue box.



Click OK, and Minitab will generate the output as shown below.

One-way ANOVA: Plant X, Plant Y, Plant Z

Method

Null hypothesis All means are equal
Alternative hypothesis At least one mean is different
Significance level $\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Factor	3	Plant X, Plant Y, Plant Z

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	2	516.0	258.00	9.00	0.003
Error	15	430.0	28.67		
Total	17	946.0			

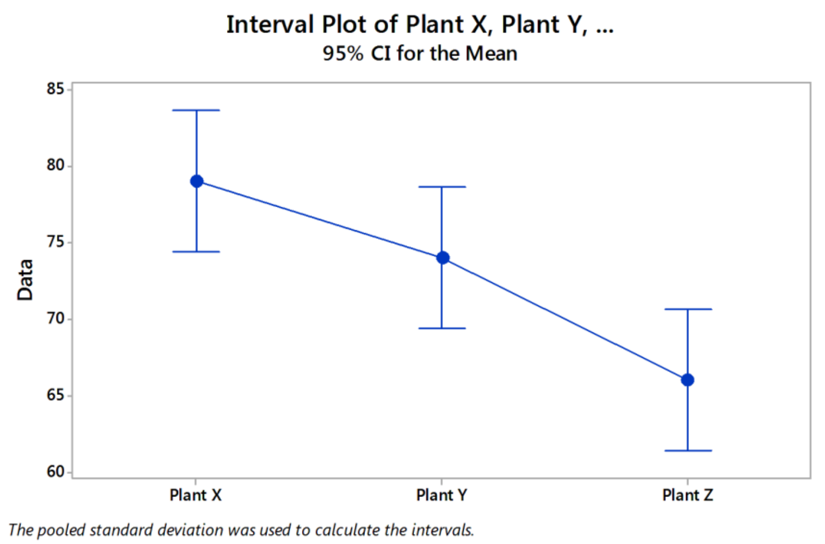
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.35413	54.55%	48.48%	34.55%

Means

Factor	N	Mean	StDev	95% CI
Plant X	6	79.00	5.83	(74.34, 83.66)
Plant Y	6	74.00	4.47	(69.34, 78.66)
Plant Z	6	66.00	5.66	(61.34, 70.66)

Pooled StDev = 5. 35413



Interpretation of Output:

Since the *p*-value = 0.003 < 0.05, we reject the null hypothesis at a 5% significance level. Therefore, the waste produced from 3 plants differs at a 5% significance level.

The *p*-value is available in the ANOVA (Analysis of Variance) section of the output. Also note that the ANOVA table is the same as that which we calculated manually in chapter 4.

A.3.7.2 Two-way ANOVA

Example of Two-way ANOVA with Replication

- A hotel is using two factors (price and city) to determine their significance on the customer satisfaction score. For each price point (Low, Medium, and High) the researcher samples each city six times (therefore, with replication). The data in the table are the customer satisfaction (CSAT) scores on a 100-point scale. The researcher is interested in understanding whether price and city had a significant effect on the satisfaction scores and whether the interaction of the two factors had a significant effect.
  Note: Interaction is a condition where a factor depends on the level of another factor to produce its effects on the response variable.

Hotel Discount - Social Media Campaign (test city & pricing as factors)				
City	Low (20)	Med (30)	High (40)	
London	77	74	49	
London	71	76	54	
London	68	69	56	
London	90	66	59	
London	72	66	55	
London	81	74	67	
Madrid	90	89	85	
Madrid	78	64	60	
Madrid	75	70	65	
Madrid	89	79	65	
Madrid	94	89	59	
Madrid	100	96	83	
Moscow	88	73	61	
Moscow	71	43	40	
Moscow	72	48	42	
Moscow	80	49	39	
Moscow	94	84	65	
Moscow	86	60	50	
San Francisco	100	92	81	
San Francisco	76	56	45	
San Francisco	87	51	39	
San Francisco	76	57	43	
San Francisco	80	66	53	
San Francisco	109	100	87	

Formulation of Hypothesis:

H_o: Row population means are same (i.e., $\mu_{London} = \mu_{Madrid} = \mu_{Moscow} = \mu_{SF}$)

H_1: At least one pair of means are different (i.e., whether city had a significant effect on CSAT score)

H_o: Column population means are same (i.e., $\mu_{Low} = \mu_{Med} = \mu_{High}$)

H_1: At least one pair of means are different (i.e., whether price had a significant effect on CSAT score)

H_o: There is no interaction between city and price

H_1: There is an interaction between city and price

Enter the given data in the Minitab worksheet as shown in the following table:

Response	Price Point	City
77	Low	London
71	Low	London
68	Low	London
90	Low	London
72	Low	London
81	Low	London
90	Low	Madrid
78	Low	Madrid
75	Low	Madrid
89	Low	Madrid
94	Low	Madrid
100	Low	Madrid
88	Low	Moscow
71	Low	Moscow
72	Low	Moscow
80	Low	Moscow
94	Low	Moscow
86	Low	Moscow
100	Low	San Francisco
76	Low	San Francisco
87	Low	San Francisco
76	Low	San Francisco
80	Low	San Francisco
109	Low	San Francisco
74	Med	London
76	Med	London
69	Med	London

66	Med	London
66	Med	London
74	Med	London
89	Med	Madrid
64	Med	Madrid
70	Med	Madrid
79	Med	Madrid
89	Med	Madrid
96	Med	Madrid
73	Med	Moscow
43	Med	Moscow
48	Med	Moscow
49	Med	Moscow
84	Med	Moscow
60	Med	Moscow
92	Med	San Francisco
56	Med	San Francisco
51	Med	San Francisco
57	Med	San Francisco
66	Med	San Francisco
100	Med	San Francisco
49	High	London
54	High	London
56	High	London
59	High	London
55	High	London
67	High	London
85	High	Madrid
60	High	Madrid
65	High	Madrid
65	High	Madrid
59	High	Madrid
83	High	Madrid
61	High	Moscow
40	High	Moscow
42	High	Moscow
39	High	Moscow
65	High	Moscow
50	High	Moscow
81	High	San Francisco
45	High	San Francisco

39	High	San Francisco
43	High	San Francisco
53	High	San Francisco
87	High	San Francisco

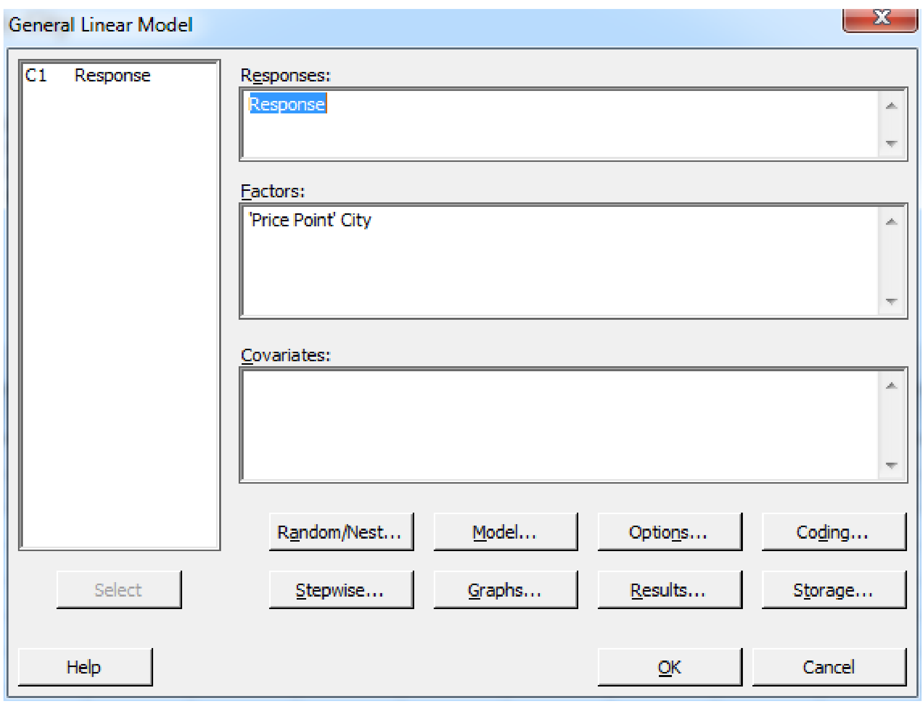
Table A.3-4: Data to be Entered in Minitab

How to perform a Two-way ANOVA in Minitab:

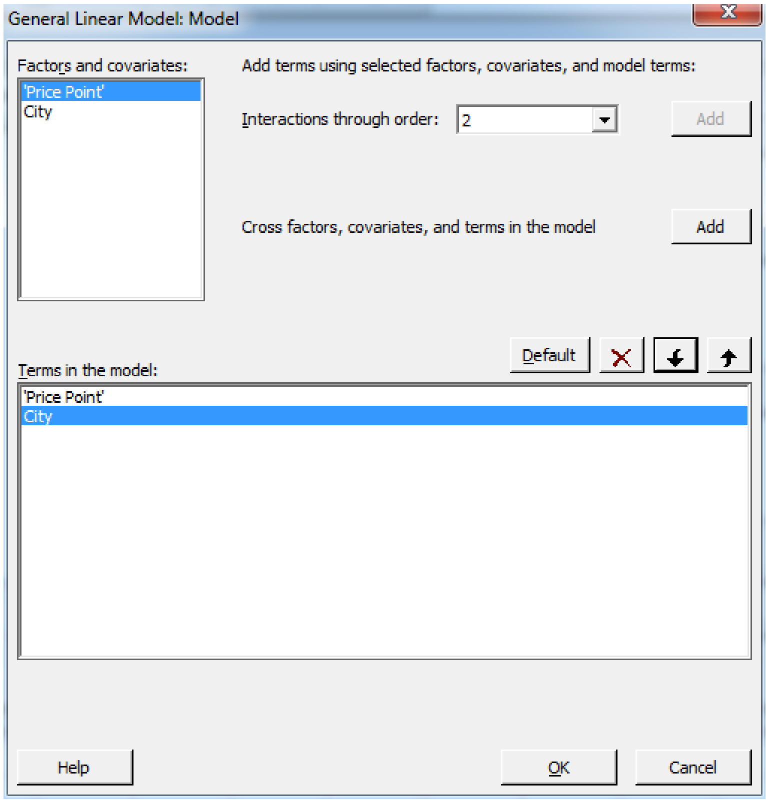
From the Stat menu:

→ Select "ANOVA" → "General Linear Model" → "Fit General Linear Model…"

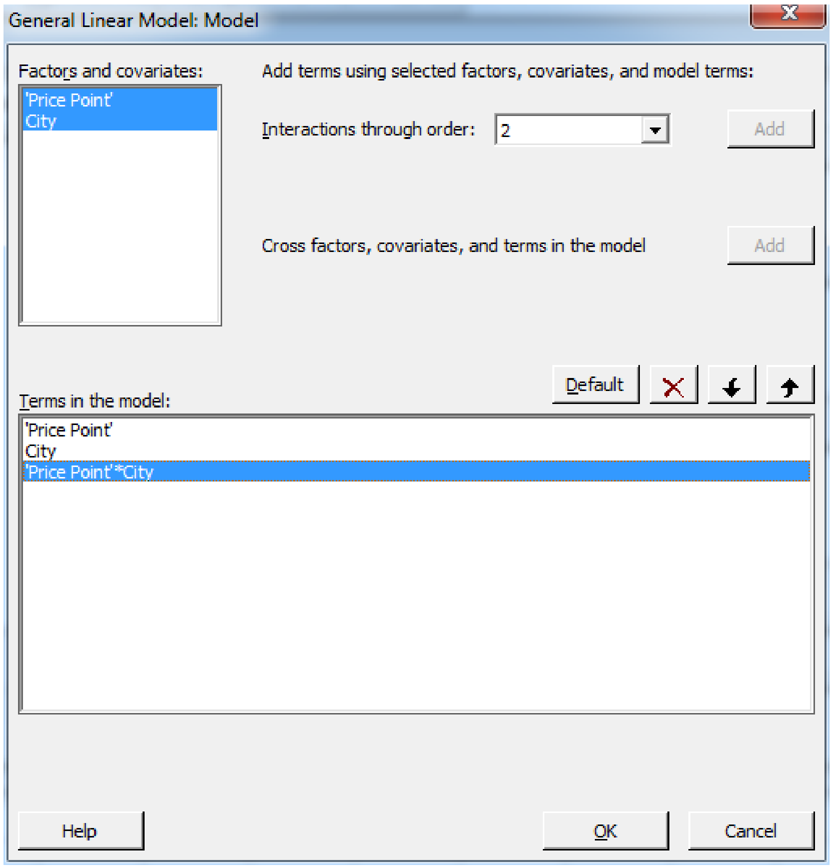
Enter the response and factors (price and city) in their respective fields as shown below in the GLM dialog box:



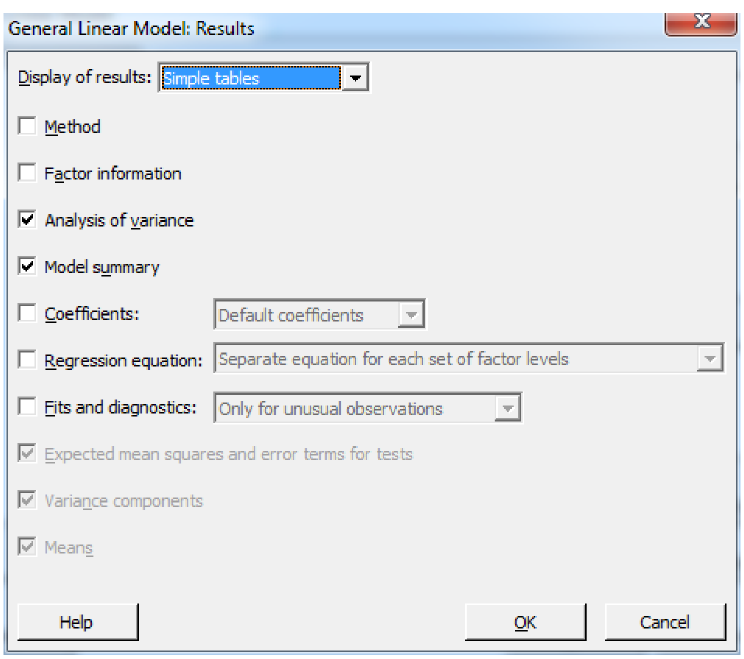
To add the interaction effect, click the Model button of the GLM dialog box. Another GLM: Model dialog box will appear as displayed below:



Select both of the factors from the left side field "Factors and covariates." The "Add" button will be activated once the factors are selected. Click the "Add" button and the interaction between the factors "Price Point*City" will be visible in the field "Terms in the Model."



Click "OK" to get back to the main GLM dialog box. Now click the "Results" button and uncheck all of the result types except "Analysis of variance" and "Model summary" to cut down on the output of unnecessary information.



Click OK → Return to the main GLM dialog box → click OK.

Minitab will produce the following output:

General Linear Model: Response versus Price Point, City

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Price Point	2	7554.1	3777.0	22.55	0.000
City	3	2447.4	815.8	4.87	0.004
Price Point*City	6	731.9	122.0	0.73	0.629
Error	60	10048.5	167.5		
Total	71	20781.9			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.9412	51.65%	42.78%	30.37%

Interpretation: The ANOVA table displays the effect of city, price, and their interaction on customer satisfaction score. From the ANOVA table, we can see that the p-values are very small (< 0.05) for both factors and therefore are considered statistically significant. But the p-value for the Interaction (Price Point*City) is high and therefore considered statistically insignificant. In other words, the effect of price on the customer satisfaction score is not dependent on city (and vice versa).

A.3.7.3 Multiple Linear Regression

Example of Multiple Linear Regression:

- The following table provides data points obtained from fifteen randomly selected trainees/delegates of a professional certification and training provider. The organization is a leading global certification provider offering certification courses in all formats (i.e., online with proctored exams as well as a traditional classroom mode). The provided data set is a random sample of collected feedback from delegates taking certification courses. Fit a linear model using the data given below, clearly stating the predictors.

Customer satisfaction score	Hours spent in class training	Hours spent in online training	Number of trial runs	Number of referrals
85	56	22	3	2
69	43	20	1	2
55	39	15	1	1
92	65	34	4	5
70	44	25	2	3
68	40	20	2	2
81	45	27	2	3
60	32	18	1	1
78	39	24	2	2
55	31	18	0	1
74	46	24	1	2
65	38	14	1	2
78	50	21	0	2
94	62	35	3	3
66	48	17	1	3

It is clear from the data that the factors affecting the output, or "Customer satisfaction scores," are hours spent in class training, hours spent in online training, number of practice tests, and number of customer referrals. These four factors are the predictors.

Since there is more than one predictor variable, we will need to conduct a multiple linear regression model.

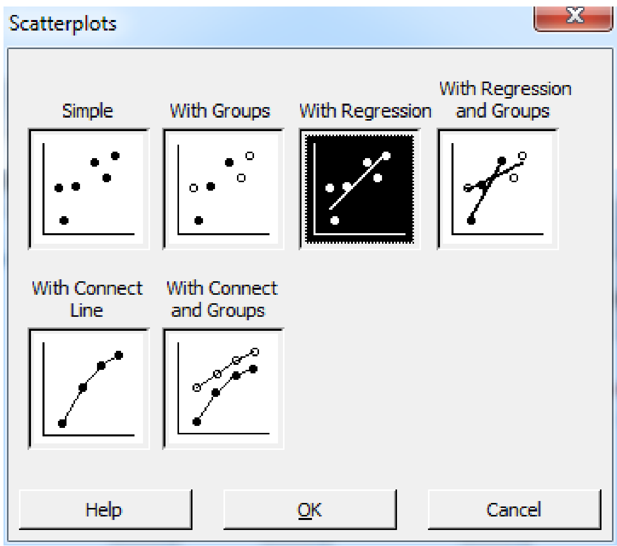
Step 1: Test for Linearity

The statistical package (Minitab) is used to determine whether the predictor variables are linear with the dependent variable.

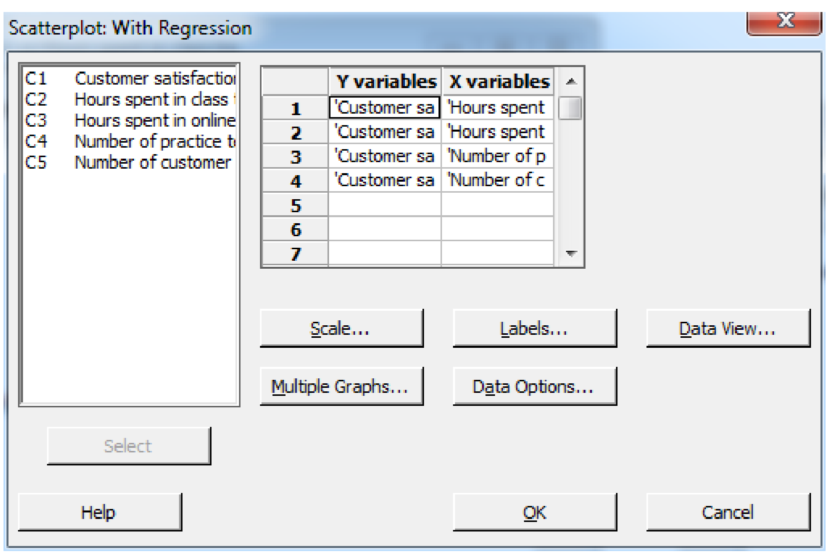
From the Stat menu:

→ Select "Graph" → "Scatterplot…"

→ Choose "With Regression…"

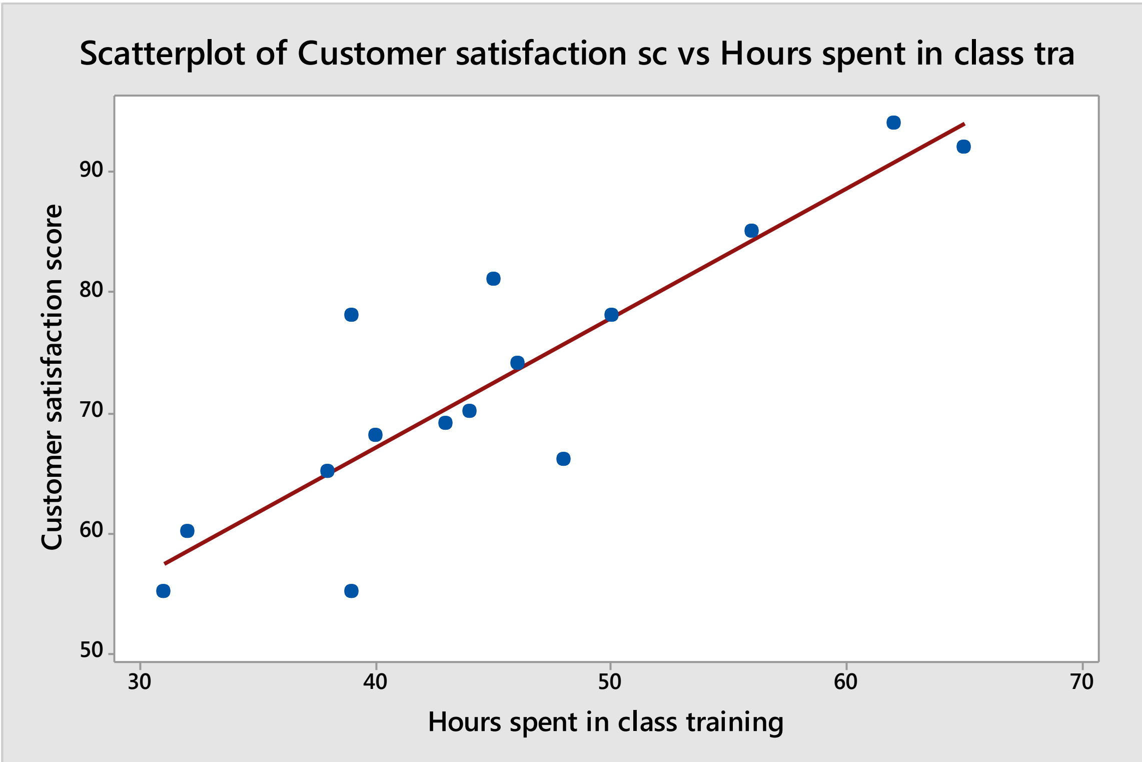


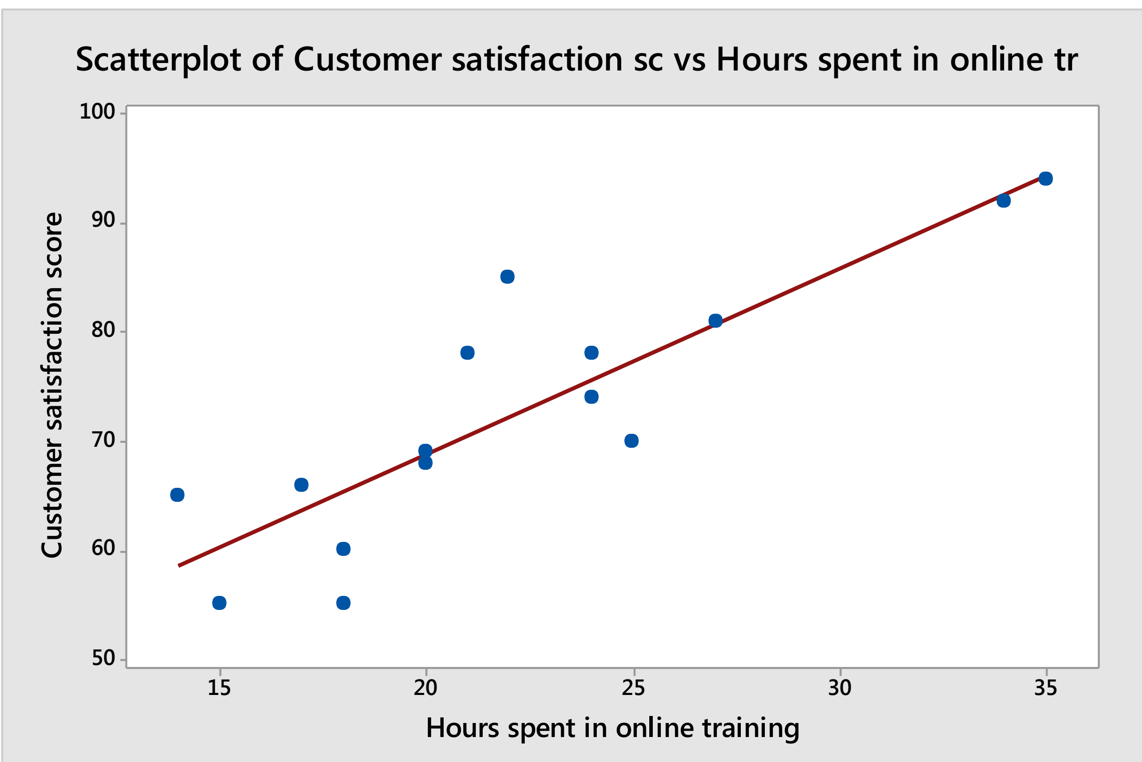
Click OK. In the dialog box that appears, select the appropriate X and Y variables as shown below:

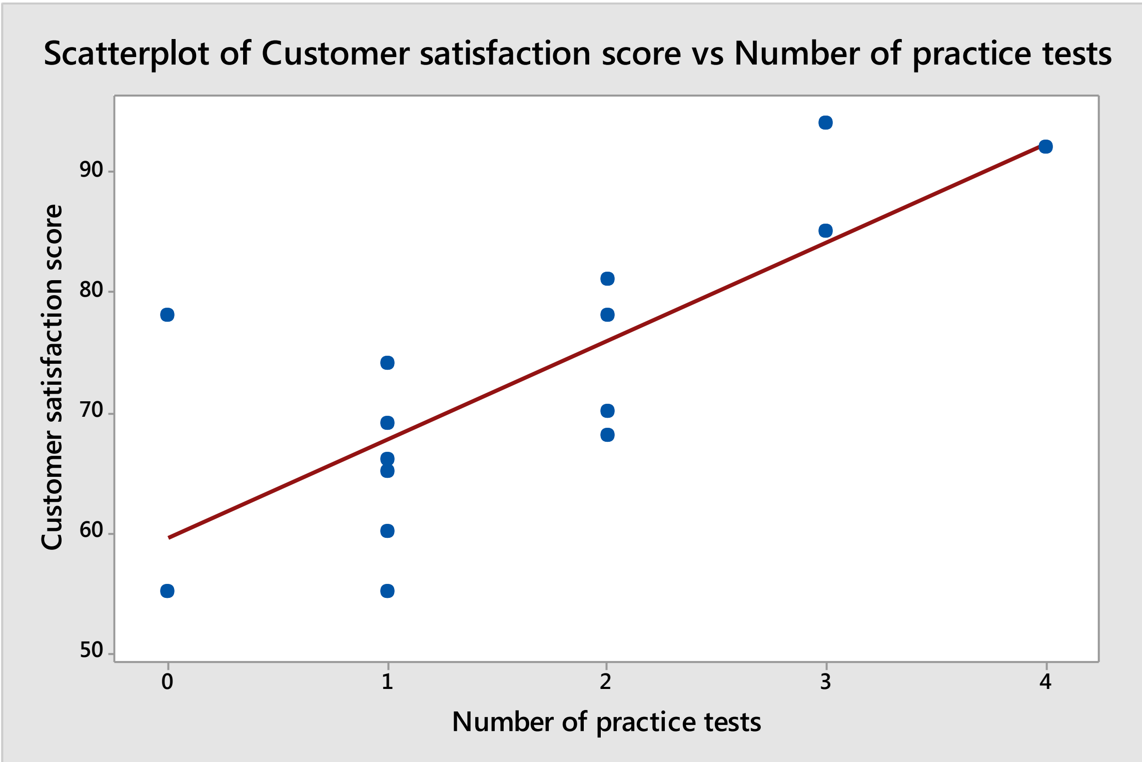


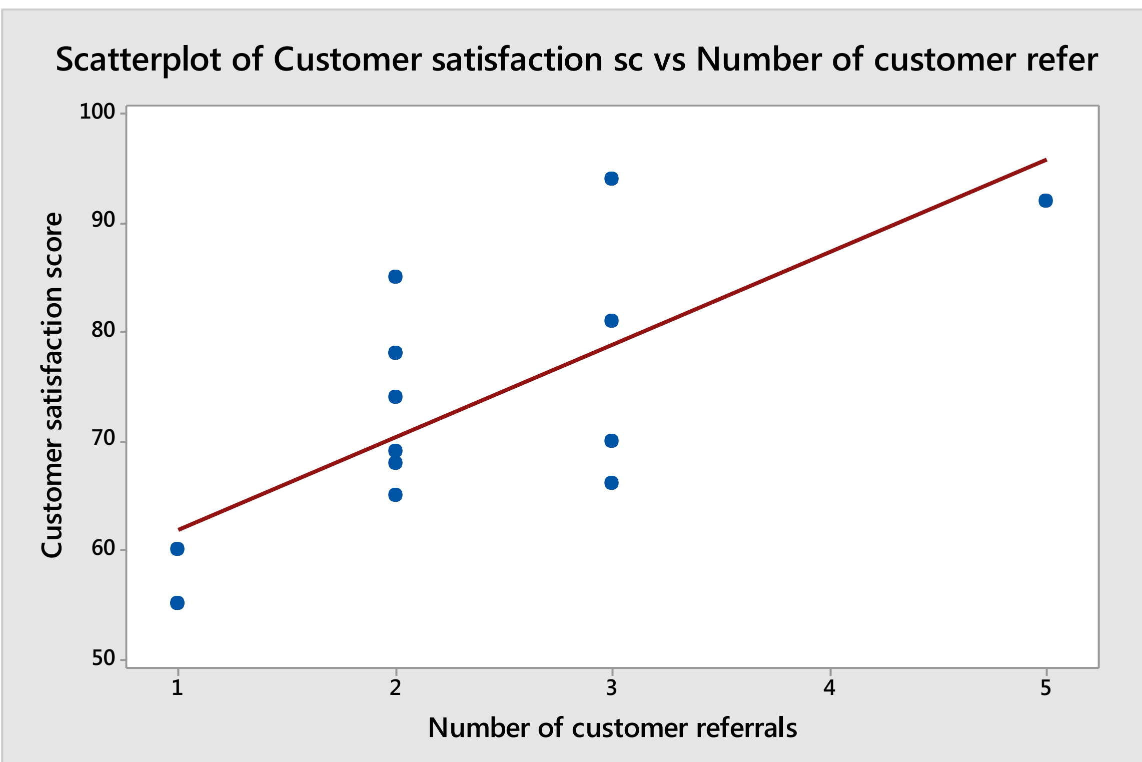
Click OK.

The outputs in the form of scatterplots are shown below:


Scatterplot of Customer satisfaction sc vs Hours spent in class tra
Customer satisfaction score
90
80
70
60
50
30
40
50
60
70
Hours spent in class training


Scatterplot of Customer satisfaction sc vs Hours spent in online tr
Customer satisfaction score
100
90
80
70
60
50
15
20
25
30
35
Hours spent in online training





Conclusion: The scatterplots fitted with the regression line indicate a linear relationship between the response variable and predictors.

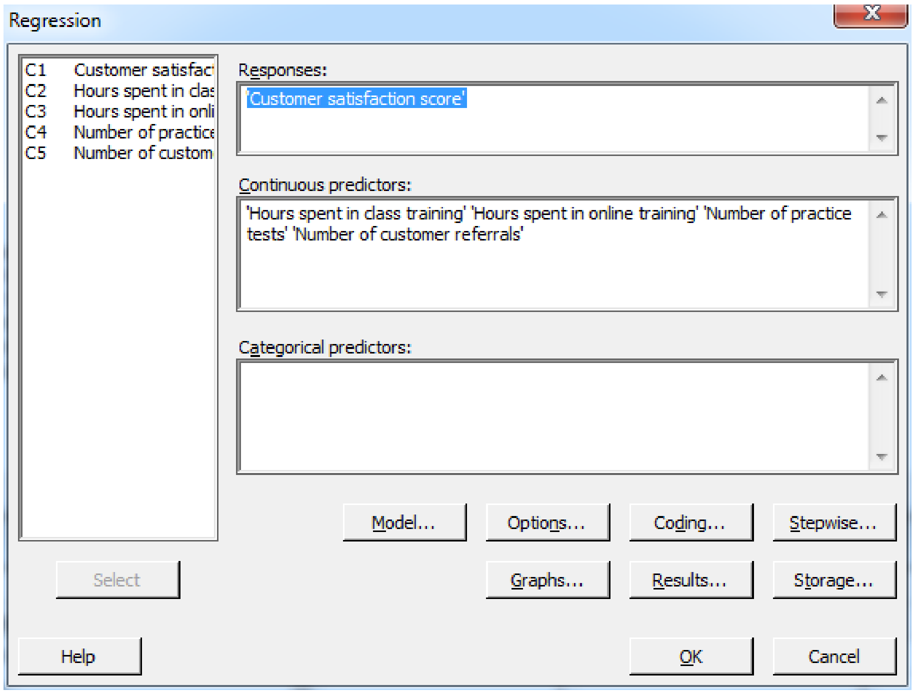
Step 2: Fit a multiple linear regression model

To perform multiple linear regression in Minitab:

From the Stat menu:

→ Select "Stat" → "Regression" → "Regression" → "Fit Regression Model…"

Select response and predictor variables as shown below:



Click OK.

The statistical package produces the output as given below:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	1790.19	447.549	17.96	0.000
Hours spent in class training	1	182.78	182.782	7.34	0.022
Hours spent in online training	1	141.69	141.686	5.69	0.038
Number of practice tests	1	8.34	8.343	0.33	0.576
Number of customer referrals	1	9.67	9.672	0.39	0.547
Error	10	249.14	24.914		
Total	14	2039.33			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.99138	87.78%	82.90%	67.57%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	23.62	7.86	3.00	0.013	
Hours spent in class training	0.665	0.246	2.71	0.022	3.29
Hours spent in online training	0.915	0.384	2.38	0.038	3.15
Number of practice tests	1.14	1.97	0.58	0.576	2.74
Number of customer referrals	-1.42	2.28	-0.62	0.547	3.10

Regression Equation

Customer satisfaction score obtained = 23.62
+ 0.665 Hours spent in class training
+ 0.915 Hours spent in online training
+ 1.14 Number of practice tests
- 1.42 Number of customer referrals

Interpretation:

There are four parts in the output. The first part, Analysis of Variance (ANOVA), provides the significance of each independent variable (predictor). The second part, the model summary, provides a measure of association (R^2 and adjusted R^2). The third part, coefficients, provides parameter estimates. The fourth part, regression equation, provides the linear model as the equation $Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + b$, where X_1, X_2, X_3 and X_4 are the predictors and Y is the output or response variable.

Value of R^2 is 87.78% (which is reasonably high). Also, the value of R^2 adjusted is close to R^2.

Test for significance of the overall model (f test) – The ANOVA table (the first part of the output) gives the F-value of the regression model, 17.96 (high). Also, the corresponding p-value is almost equal to zero. Therefore, the overall fit is good.

Test for multicolinearity – Check for the VIF value for each predictor variable. Since none are greater than 5, we can conclude that there does not exist any multicolinearity.

Test for significance of each predictor (t test) – The coefficients table gives t-value as well as p-value for each predictor. Note that the t-values for “number of practice tests” and “number of customer referrals” are low and that the corresponding p-value is high (greater than 0.05). Therefore, we can conclude that these two predictors are insignificant. Once insignificant predictors are identified they may be removed from the model to make the analysis more accurate.

This can be done using a technique called “Stepwise,” available in many statistical packages such as Minitab, SAS, and SPSS.

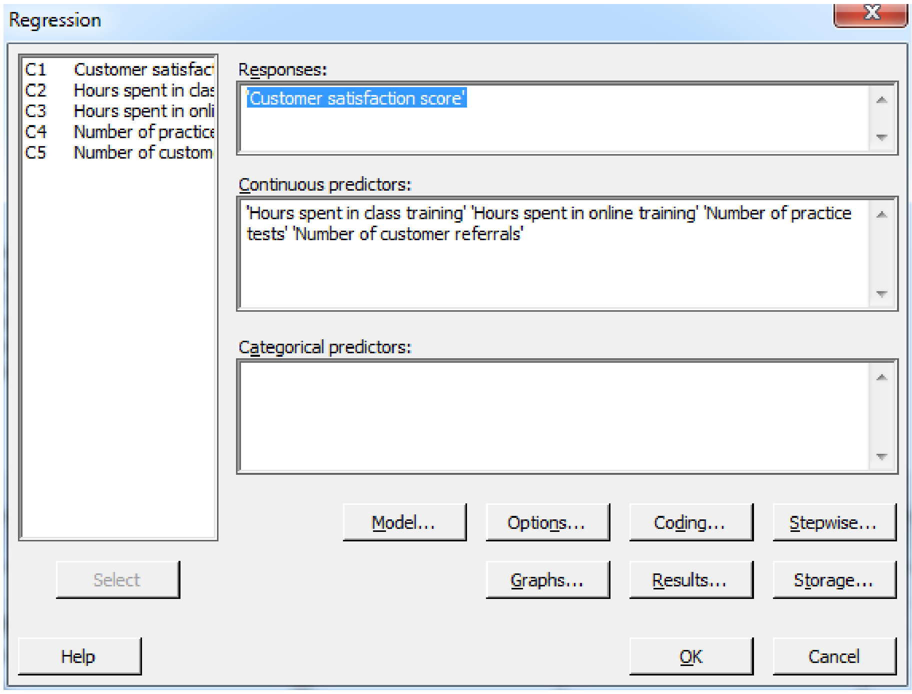
Use Stepwise to reduce insignificant variables.

To perform the “Stepwise” method as part of the multiple linear regression in Minitab:

From the Stat menu:

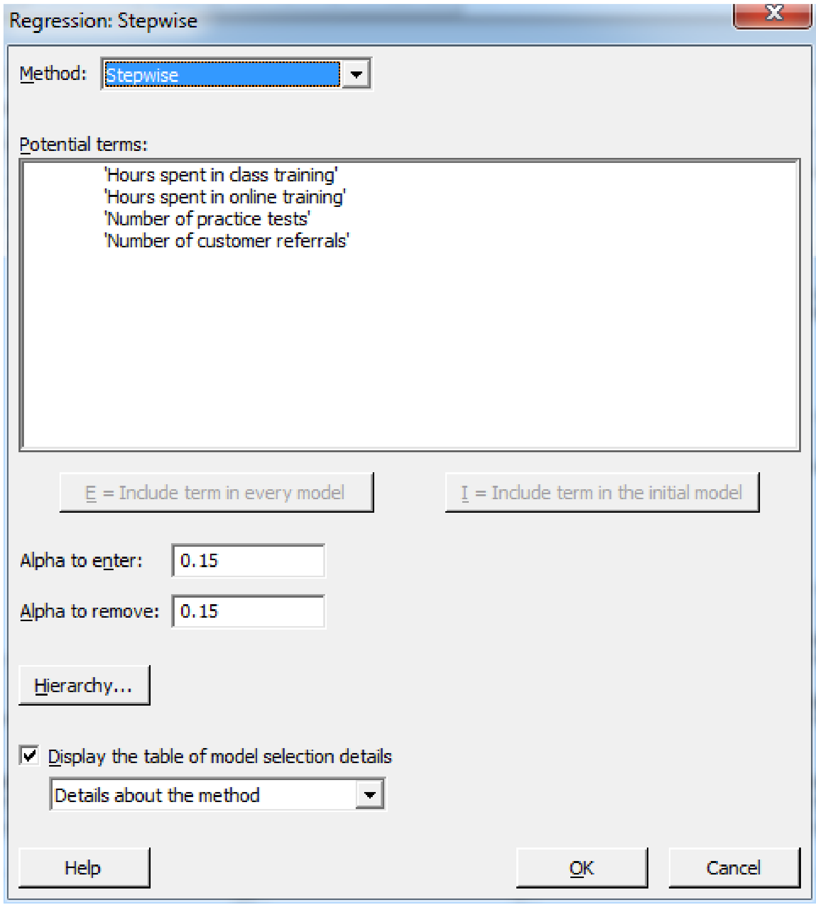
→ Select “Stat” → “Regression” → “Regression” → “Fit Regression Model…”

Select response and predictor variables as shown below:



Click “Stepwise…”

Select the method “Stepwise” in the drop down box as shown below:



Click OK to get the output as shown below:

Regression Analysis: Customer sat versus Hours spent, Hours spent, Number of practice etc.

Stepwise Selection of Terms

α to enter = 0.15, α to remove = 0.15

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1775.6	887.79	40.39	0.000
Hours spent in class training	1	230.8	230.81	10.50	0.007
Hours spent in online training	1	195.8	195.78	8.91	0.011
Error	12	263.8	21.98		
Total	14	2039.3			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
4.68829	87.07%	84.91%	78.62%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	23.25	5.88	3.96	0.002	
Hours spent in class training	0.634	0.196	3.24	0.007	2.37
Hours spent in online training	0.933	0.313	2.98	0.011	2.37

Regression Equation

Customer satisfaction score = 23.25 + 0.634 Hours spent in class training
+ 0.933 Hours spent in online training

Interpretation:

As you can see, the two insignificant predictors have been removed in this model. Also the F-value of the regression model was improved (increased to 40.30 from 17.96 in the earlier model). The difference between R^2 and adjusted R^2 is also reduced. Therefore, the improved linear model is "Customer satisfaction score obtained = 23.25 + 0.634 (Hours spent in class training) + 0.933 (Hours spent in online training)."

A.3.7.4. Factor Analysis

Example of Factor Analysis:

- An organization is involved in publishing college level textbooks on various subjects. It has employed a number of sales people in this regard and is interested in identifying the factors of a successful sales person. The HR department maintains an information database of the sales people. The following table contains data pertaining to ten sales people in seven variables: height, weight, age, educational qualification, number of children, size of household, and IQ level. The organization would like to reduce these seven variables to three factors for better insight into the criteria for predicting success in the sales role.

Sales Person	Height	Weight	Education	Age	No. of Children	Size of Household	IQ
1	69	168	14	33	1	3	112
2	71	175	16	24	0	1	116
3	72	182	12	35	2	4	109
4	68	165	11	42	3	5	91
5	73	190	13	30	1	2	115
6	68	162	16	33	1	3	117
7	70	173	12	45	4	6	120
8	71	179	13	25	0	2	93
9	66	150	10	39	2	4	100
10	76	210	16	26	0	1	110

We have used an advanced statistical package Minitab to solve this problem. The same can be done using other advanced statistical packages such as SPSS and SAS. All these packages have built-in programs that run at the backend to provide the output.

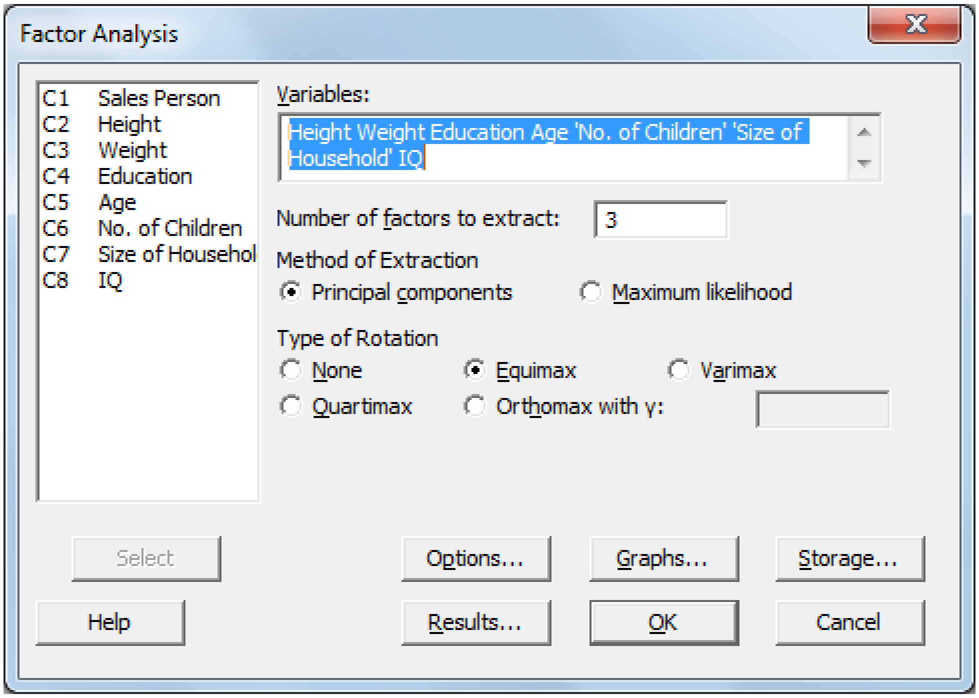
To perform a factor analysis in Minitab:

From the Stat menu

→ Select “Multivariate”

→ Choose “Factor Analysis…”

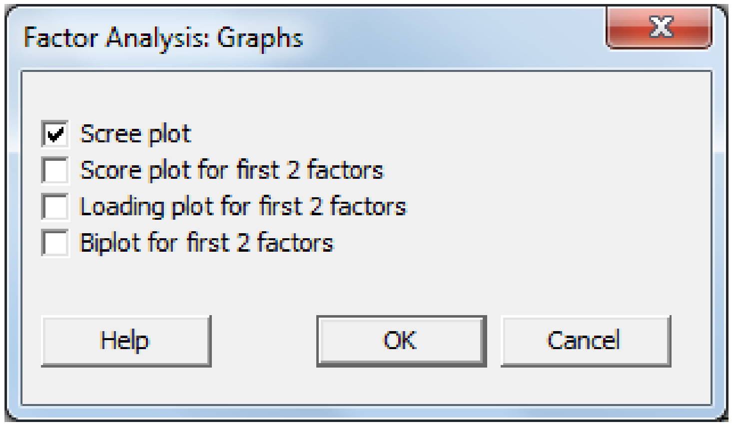
The Factor Analysis dialog box is shown below.



1. Select all seven variables as the variables to be reduced.
2. Enter 3 as the "Number of factors to extract." Number of factors can be determined using Eigen value or Scree-plot method. In the Eigen value method, retain the factors with Eigen values > 1. In the Scree-plot method, the true number of factors is indicated by the point where the Scree begins—Elbow criterion.
3. Select the Method of Extraction as "Principal Components." Principal Components Analysis is used when the primary concern is to determine the minimum number of factors that account for the maximum variance in the data, for use in subsequent analysis.
4. Select the Type of Rotation as "Equimax." The following are some rotation methods provided by Minitab:
   - Varimax: An orthogonal rotation method that minimizes the number of variables that have high loadings on each factor.
   - Quartimax: A rotation method that minimizes the number of factors needed to explain each variable.
   - Equimax: A combination of Varimax and Quartimax method. Both the number of variables that load highly on a factor and the number of factors needed to explain a variable are minimized.

   It is common practice to use the Equimax rotation method for factor analysis.

5. Click on “Graphs…” and select “Scree plot” in the dialog box and then click OK.



6. Click OK in the main dialog box to get the Minitab output with Scree plot as given below.

Factor Analysis: Height, Weight, Education, Age, No. of Children, Size of Household, IQ

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Height	0.776	-0.432	-0.457	0.998
Weight	0.750	-0.423	-0.504	0.995
Education	0.816	-0.136	0.450	0.887
Age	-0.932	-0.326	-0.035	0.977
No. of Children	-0.885	-0.439	-0.112	0.988
Size of Household	-0.938	-0.296	-0.071	0.972
IQ	0.283	-0.728	0.591	0.960
Variance	4.4418	1.3006	1.0344	6.7768

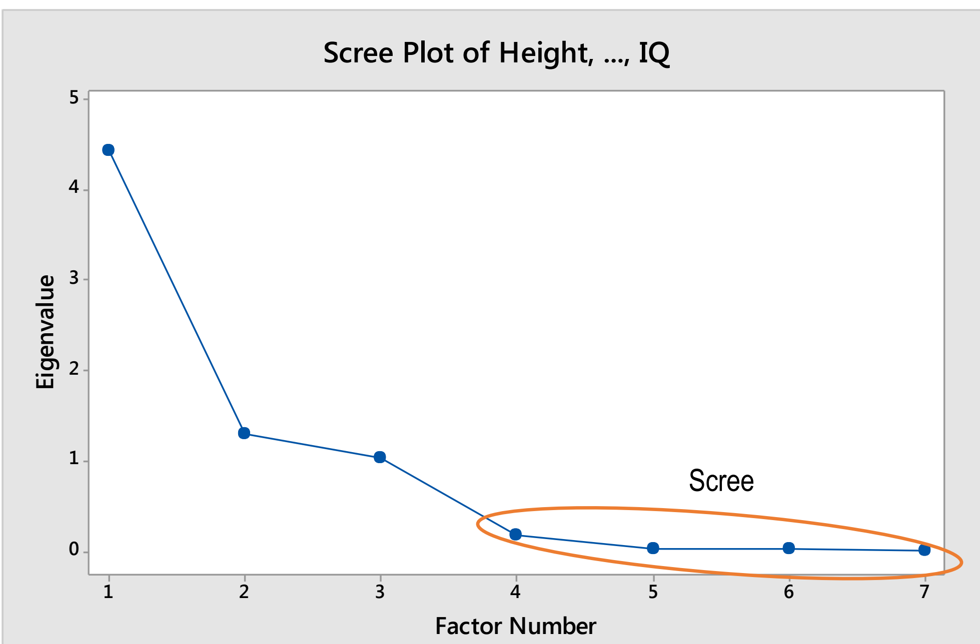
Rotated Factor Loadings and Communalities
Equimax Rotation

Variable	Factor1	Factor2	Factor3	Communality
Height	-0.249	0.951	-0.175	0.998
Weight	-0.223	0.964	-0.129	0.995
Education	-0.664	0.221	-0.631	0.887
Age	0.925	-0.341	0.067	0.977
No. of Children	0.971	-0.208	0.035	0.988
Size of Household	0.920	-0.335	0.114	0.972
IQ	0.062	0.115	-0.971	0.960
Variance	3.2025	2.1674	1.4069	6.7768
% Var	0.457	0.310	0.201	0.968

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3
Height	0.159	0.556	0.045
Weight	0.170	0.579	0.083
Education	-0.185	-0.13	-0.425
Age	0.317	0.027	-0.083
No. of Children	0.376	0.124	-0.090
Size of Household	0.313	0.038	-0.043
IQ	0.141	-0.07	-0.787

Scree Plot of Height, ... , IQ



Interpretation of Output

The Scree plot suggests that the 3 factor solution contributes the larger part of the data because the Eigen values of factors 4 to 7 are almost zero with little change in the variation contributed.

From Rotated Factor Matrix:

- Factor 1 is highly loaded on number of children, age of the sales person, and size of the household. Since loading on number of children is more than other variables, it better represents this group.
- Factor 2 is highly loaded on weight and height of the sales person, which represents the physical fitness of the person. Since loading on weight is more as compared to height, weight is a better representative of these two variables.
- Factor 3 is highly loaded on IQ level. Here is an interesting fact to be noticed: educational qualification is highly loaded with both Factor 1 and 3 making it complicated to associate with any of the factors. In reality, the educational qualification of a person is closer to the mental ability that a person possesses, and thus can be grouped with IQ level, making IQ level the representative of the third component or factor.

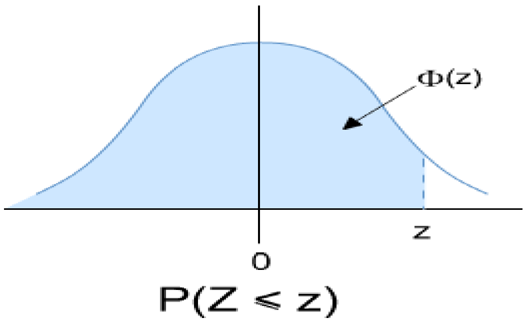
In this case, a larger dataset is reduced to a relatively smaller dataset (seven variables are reduced to three factors, each factor being represented by one initial variable) without losing any critical information.

A.3.8 Statistical Tables

Standard Normal Table

The following table represents the percentage of the shaded region in the chart below, denoted as $\Phi(z)$, under the standard normal curve corresponding to the z value. Therefore, $Z_{\frac{\alpha}{2}}$ is the value of z corresponding to $\Phi(z) = 0.975$ because $Z_{\frac{\alpha}{2}}$ is the upper 2.5% point of the standard normal distribution. In other words, it is the value of the standard normal variable z, which covers 97.5% of the area on its left and 100 – 97.5 = 2.5% of the area on its right.

z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)
1.20	0.88493	1.90	0.97128	2.60	0.99534	3.30	0.99952	1.55	0.93943	2.25	0.98778	2.95	0.99841	3.65	0.99987
1.21	0.88686	1.91	0.97193	2.61	0.99547	3.31	0.99953	1.56	0.94062	2.26	0.98809	2.96	0.99846	3.66	0.99987
1.22	0.88877	1.92	0.97257	2.62	0.99560	3.32	0.99955	1.57	0.94179	2.27	0.98840	2.97	0.99851	3.67	0.99988
1.23	0.89065	1.93	0.97320	2.63	0.99573	3.33	0.99957	1.58	0.94295	2.28	0.98870	2.98	0.99856	3.68	0.99988
1.24	0.89251	1.94	0.97381	2.64	0.99585	3.34	0.99958	1.59	0.94408	2.29	0.98899	2.99	0.99861	3.69	0.99989
1.25	0.89435	1.95	0.97441	2.65	0.99598	3.35	0.99960	1.60	0.94520	2.30	0.98928	3.00	0.99865	3.70	0.99989
1.26	0.89617	1.96	0.97500	2.66	0.99609	3.36	0.99961	1.61	0.94630	2.31	0.98956	3.01	0.99869	3.71	0.99990
1.27	0.89796	1.97	0.97558	2.67	0.99621	3.37	0.99962	1.62	0.94738	2.32	0.98983	3.02	0.99874	3.72	0.99990
1.28	0.89973	1.98	0.97615	2.68	0.99632	3.38	0.99964	1.63	0.94845	2.33	0.99010	3.03	0.99878	3.73	0.99990
1.29	0.90147	1.99	0.97670	2.69	0.99643	3.39	0.99965	1.64	0.94950	2.34	0.99036	3.04	0.99882	3.74	0.99991
1.30	0.90320	2.00	0.97725	2.70	0.99653	3.40	0.99966	1.65	0.95053	2.35	0.99061	3.05	0.99886	3.75	0.99991
1.31	0.90490	2.01	0.97778	2.71	0.99664	3.41	0.99968	1.66	0.95154	2.36	0.99086	3.06	0.99889	3.76	0.99992
1.32	0.90658	2.02	0.97831	2.72	0.99674	3.42	0.99969	1.67	0.95254	2.37	0.99111	3.07	0.99893	3.77	0.99992
1.33	0.90824	2.03	0.97882	2.73	0.99683	3.43	0.99970	1.68	0.95352	2.38	0.99134	3.08	0.99896	3.78	0.99992
1.34	0.90988	2.04	0.97932	2.74	0.99693	3.44	0.99971	1.69	0.95449	2.39	0.99158	3.09	0.99900	3.79	0.99992
1.35	0.91149	2.05	0.97982	2.75	0.99702	3.45	0.99972	1.70	0.95543	2.40	0.99180	3.10	0.99903	3.80	0.99993
1.36	0.91309	2.06	0.98030	2.76	0.99711	3.46	0.99973	1.71	0.95637	2.41	0.99202	3.11	0.99906	3.81	0.99993
1.37	0.91466	2.07	0.98077	2.77	0.99720	3.47	0.99974	1.72	0.95728	2.42	0.99224	3.12	0.99910	3.82	0.99993
1.38	0.91621	2.08	0.98124	2.78	0.99728	3.48	0.99975	1.73	0.95818	2.43	0.99245	3.13	0.99913	3.83	0.99994
1.39	0.91774	2.09	0.98169	2.79	0.99736	3.49	0.99976	1.74	0.95907	2.44	0.99266	3.14	0.99916	3.84	0.99994
1.40	0.91924	2.10	0.98214	2.80	0.99744	3.50	0.99977	1.75	0.95994	2.45	0.99286	3.15	0.99918	3.85	0.99994
1.41	0.92073	2.11	0.98257	2.81	0.99752	3.51	0.99978	1.76	0.96080	2.46	0.99305	3.16	0.99921	3.86	0.99994
1.42	0.92220	2.12	0.98300	2.82	0.99760	3.52	0.99978	1.77	0.96164	2.47	0.99324	3.17	0.99924	3.87	0.99995
1.43	0.92364	2.13	0.98341	2.83	0.99767	3.53	0.99979	1.78	0.96246	2.48	0.99343	3.18	0.99926	3.88	0.99995
1.44	0.92507	2.14	0.98382	2.84	0.99774	3.54	0.99980	1.79	0.96327	2.49	0.99361	3.19	0.99929	3.89	0.99995
1.45	0.92647	2.15	0.98422	2.85	0.99781	3.55	0.99981	1.80	0.96407	2.50	0.99379	3.20	0.99931	3.90	0.99995
1.46	0.92785	2.16	0.98461	2.86	0.99788	3.56	0.99981	1.81	0.96485	2.51	0.99396	3.21	0.99934	3.91	0.99995
1.47	0.92922	2.17	0.98500	2.87	0.99795	3.57	0.99982	1.82	0.96562	2.52	0.99413	3.22	0.99936	3.92	0.99996
1.48	0.93056	2.18	0.98537	2.88	0.99801	3.58	0.99983	1.83	0.96638	2.53	0.99430	3.23	0.99938	3.93	0.99996
1.49	0.93189	2.19	0.98574	2.89	0.99807	3.59	0.99983	1.84	0.96712	2.54	0.99446	3.24	0.99940	3.94	0.99996
1.50	0.93319	2.20	0.98610	2.90	0.99813	3.60	0.99984	1.85	0.96784	2.55	0.99461	3.25	0.99942	3.95	0.99996
1.51	0.93448	2.21	0.98645	2.91	0.99819	3.61	0.99985	1.86	0.96856	2.56	0.99477	3.26	0.99944	3.96	0.99996
1.52	0.93574	2.22	0.98679	2.92	0.99825	3.62	0.99985	1.87	0.96926	2.57	0.99492	3.27	0.99946	3.97	0.99996
1.53	0.93699	2.23	0.98713	2.93	0.99831	3.63	0.99986	1.88	0.96995	2.58	0.99506	3.28	0.99948	3.98	0.99997
1.54	0.93822	2.24	0.98745	2.94	0.99836	3.64	0.99986	1.89	0.97062	2.59	0.99520	3.29	0.99950	3.99	0.99997



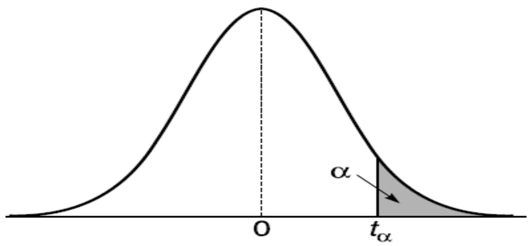
t-Distribution

The t-table displays t-values corresponding to the significance level α (column wise) and degrees of freedom (row wise). For example, the critical value of t-distribution with 11 degrees of freedom corresponding to a 5% significance level is 2.201. This is derived by taking an upper tail area of 2.5%. Since t-distribution is symmetric around the mean, the critical region (i.e., 5% of the total area under the curve), is equally distributed on either side of the mean—2.5% as upper tail area and 2.5% as lower tail area. Therefore, an upper tail area of 2.5% is taken to derive the critical value.

Upper tail area(α)

Degrees of Freedom

df	0.250	0.100	0.050	0.025	0.010	0.005
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.897	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.625	2.977
15	0.691	1.341	1.753	2.132	2.603	2.947
16	0.690	1.337	1.746	2.120	2.584	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.540	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.320	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
∞	0.675	1.282	1.645	1.960	2.326	2.576



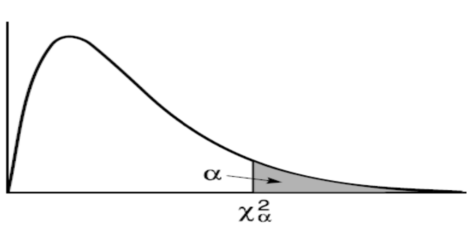
Chi-square Distribution

The Chi-square table displays chi-square values corresponding to the significance level (column wise) and degrees of freedom (row wise). For example, to determine the critical value of Chi-square distribution with 5 degrees of freedom corresponding to 5% significance level, the value corresponding to the α value of 0.05 and a df value of 5 should be examined. From the table provided, the critical value is 11.071.

Upper tail area (α)

Degrees of Freedom

df	0.995	0.990	0.975	0.950	0.900	0.750	0.250	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	14.845	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	9.299	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	18.245	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	11.912	19.369	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	12.792	20.489	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	13.675	21.605	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	14.562	22.718	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	15.452	23.828	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	16.344	24.935	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	17.240	26.039	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	18.137	27.141	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	19.037	28.241	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	19.939	29.339	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	20.843	30.435	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	21.749	31.528	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	22.657	32.620	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	23.567	33.711	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	24.478	34.800	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	25.390	35.887	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	26.304	36.973	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	27.219	38.058	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	28.136	39.141	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	29.054	40.223	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	29.973	41.304	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	30.893	42.383	48.363	52.192	55.668	59.892	62.883
38	19.289	20.691	22.878	24.884	27.343	31.815	43.462	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	32.737	44.539	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	33.660	45.616	51.805	55.758	59.342	63.691	66.766
41	21.421	22.906	25.215	27.326	29.907	34.585	46.692	52.949	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	30.765	35.510	47.766	54.090	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	31.625	36.436	48.840	55.230	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	32.487	37.363	49.913	56.369	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	33.350	38.291	50.985	57.505	61.656	65.410	69.957	73.166
46	25.041	26.657	29.160	31.439	34.215	39.220	52.056	58.641	62.830	66.617	71.201	74.437
47	25.775	27.416	29.956	32.268	35.081	40.149	53.127	59.774	64.001	67.821	72.443	75.704
48	26.511	28.177	30.755	33.098	35.949	41.079	54.196	60.907	65.171	69.023	73.683	76.969
49	27.249	28.941	31.555	33.930	36.818	42.010	55.265	62.038	66.339	70.222	74.919	78.231
50	27.991	29.707	32.357	34.764	37.689	42.942	56.334	63.167	67.505	71.420	76.154	79.490
51	28.735	30.475	33.162	35.600	38.560	43.874	57.401	64.295	68.669	72.616	77.386	80.747
52	29.481	31.246	33.968	36.437	39.433	44.808	58.468	65.422	69.832	73.810	78.616	82.001
53	30.230	32.018	34.776	37.276	40.308	45.741	59.534	66.548	70.993	75.002	79.843	83.253
54	30.981	32.793	35.586	38.116	41.183	46.676	60.600	67.673	72.153	76.192	81.069	84.502
55	31.735	33.570	36.398	38.958	42.060	47.610	61.665	68.796	73.311	77.380	82.292	85.749
56	32.490	34.350	37.212	39.801	42.937	48.546	62.729	69.919	74.468	78.567	83.513	86.994
57	33.248	35.131	38.027	40.646	43.816	49.482	63.793	71.040	75.624	79.752	84.733	88.236
58	34.008	35.913	38.844	41.492	44.696	50.419	64.857	72.160	76.778	80.936	85.950	89.477
59	34.770	36.698	39.662	42.339	45.577	51.356	65.919	73.279	77.931	82.117	87.166	90.715
60	35.534	37.485	40.482	43.188	46.459	52.294	66.981	74.397	79.082	83.298	88.379	91.952



F Distribution

To find the critical value of F from the F table, significance level (α), degrees of freedom associated with a factor, and degrees of freedom associated with error are required. The significance level determines which F table is to be used. For example, if significance level (α) = 0.025, then F.975 table should be used; if significance level (α) = 0.05, then F.95 table should be used. Numerator df is the degrees of freedom associated with a factor and denominator df is the degrees of freedom associated with error. Each F table (F.975, F.95) displays the F-values corresponding to the numerator (columns) and denominator (rows) df.

F distribution $F_{.95}$

	Numerator Degrees of Freedom																		
df (Denominator Degrees of Freedom)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90	243.90	245.90	248.00	249.10	250.10	251.10	252.20	253.30	254.30
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F Distribution $F_{.975}$

Numerator Degrees of Freedom

Denominator Degrees of Freedom

df	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.80	799.50	864.20	899.60	921.80	937.10	948.20	956.70	963.30	968.60	976.70	984.90	993.10	997.20	1001.00	1006.00	1010.00	1014.00	1018.00
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

APPENDIX B. AUTHORS AND REVIEWERS OF THE SMBOK® GUIDE

This appendix lists the names of those individuals who contributed to the development and production of the *SMBOK® Guide*.

SMstudy® is grateful to all these individuals for their continuous support and acknowledges their contributions towards the development of the *SMBOK® Guide*.

Authors and Subject Matter Experts

Tridibesh Satpathy

Shawna Sheldon

Trent Fogg

Ruth Kim

Jaimie Rush

Gaurav Pathak

Elizabeth Airhart

Joanne Morrison

Holly Bowyer

Leslie Sloan

Lawrence Hourihan

Derek Davidson

Karen Lyncook

Melissa Lauro

James Pruitt

Nathan K. Cartmell

Pritam Sarkar

Suman Ghosh

Alexis Carnegie Dunham

Amit Pradhan

Bhanu Prasad

Santosh Bhojarajan

Carrie Ann Barrow

Daniel K Wentzel

Dr. Giridhar Thirumalai

Eduardo Barraza

Ernesto Ibarra

Ganesh Watve

Imran K Malik

Jacqueline Grant

James Paul

Javier Contreras

Joel Serrano

John Katsiris

Kim Essendrup

Micheal Olufemi Falodun

Mukul Gupta

Praveen George Jomy

Prayas Kad

Ralf Buhl

Sasan Nikookar

Warren Krams

Zimele Cele

REFERENCES

Covey, Stephen R. (1989). “The 7 Habits of Highly Effective People”. Free Press

Luft, J., Ingham, H. (1955). "The Johari Window, A Graphic Model of Interpersonal Awareness". Proceedings of the Western Training Laboratory in Group Development, University of California

Porter, M.E. (1998). “The Five Competitive Forces that Shape Strategy.” Competitive Advantage. New York: The Free Press

Beri, G.C, (2008). “Marketing Research”. 4th Ed. McGraw Hill

Webb, John. R. (2002). “Understanding & Designing Marketing Research”. London, Thomson Learning

Burns, Alvin C. and Bush, Ronald F. (2009). “Marketing Research” – Fifth Edition

Clow, Kenneth, E. and James, Karen E. (2014). “Essentials of Marketing Research – Putting Research into Practice” Sage

Malhotra, Naresh K. (2008). “Marketing Research: An Applied Orientation”. Prentice Hall

GLOSSARY

2X2 matrix

The use of a 2X2 matrix for improving decision-making and problem identification crosses many fields of study and applications from game theory to consulting. A 2X2 matrix can expand options and force the team to think muti-dimensionally. It creates focus around the problem statement to allow exploration of various types of opposition— direct, complementary, and reflexive.

5C Analysis

5C Analysis is one of the most popular and useful frameworks in understanding internal and external environments. It is an extension of the 3C Analysis that originally included Company, Customers, and Competitors. Collaborators and Climate were later added to the analysis to make it comprehensive. This integrated analysis covers the most important areas of marketing, and the insights generated can help identify the key problems and challenges facing the organization.

After-only Design

This is the simplest of all the experimental designs. The independent variable is exposed to the alteration and the result is observed after a period of time.

Alternative Hypothesis

This is the hypothesis to be accepted when the null hypothesis is not true. It is the opposite of the null hypothesis and is denoted by H1 or Ha.

Analysis of Variance (ANOVA)

The ANOVA technique is used when testing differences between two or more means of independent samples. There are two parts in ANOVA: one-way classification and two-way classification.

Annual Reports

Annual reports provide either internal or external secondary data and can be used as a good starting point. It may have information such as sales records, marketing activity, cost information, distribution channels, and customer feedback. Data is collected and organized in accordance with the business the company conducts.

Aspects of Sales and Marketing

Also referred to as “Aspects of Marketing”. The Aspects of Sales and Marketing are based on the six most common and often distinct career fields related to Sales and Marketing. They are Marketing Strategy (MS), Marketing Research (MR), Digital Marketing (DM), Corporate Sales (CS), Branding and Advertising, and Retail Marketing (RM).

Association Technique

In this type of projective technique, subjects are provided with a hint or stimulus and asked to respond with the first thing that comes to mind. Word association is a common example of the association technique.

Attitudes

Attitudes refer to a person’s feelings, convictions, or beliefs toward an object, idea, or an individual. Since attitude impacts behavior, it is of great importance to marketers. It is difficult to measure or observe behavior in all conditions, and measuring attitudes of the respondents can provide a good indication of possible behaviors.

Awareness/Knowledge

This data refers to what subjects do or do not know about an object of investigation. Information influences behavior and marketers often want to know how the behavior of customers changes with their level of awareness regarding a particular product, brand, object, or industry. This information helps marketers determine the image, experience, and feelings of consumers who are familiar with a product and make distinctions between consumers who are familiar with a particular product or industry and those who are not.

Background information

Background information is an integral part of the research design. Background information puts the research objectives into context, helping the researcher understand why certain research objectives are being pursued. The background information also gives a framework for the researcher to investigate other potential events and contributing factors or causes in addition to the defined research objectives. The background information provides hints to the researcher regarding what information he or she should be looking for and where to look for it.

Bar Charts

Bar charts, as the name suggests, consist of a series of bars of equal thickness and whose lengths represent the value of items. Bar charts show changes in a dependent variable at discrete intervals of the independent variable.

Barter System

More than a thousand years ago, when coins and other forms of money were not yet popular, the typical and most common way people procured their products or services was through the barter system—the direct exchange of goods or services without the use of money.

Before-After Design

In this design, first the dependent variable is measured, then the independent variable is altered and finally the dependent variable is measured again to understand the impact of the treatment on the test units. This design differs from the after-only design in that it has both a pretest as well as a posttest.

Before-After with Control Group

As part of the before-after design with the control group, the test group is divided into two homogenous groups—a control group and an experimental group. A pretest measurement of the dependent variable is taken for both the groups. The experimental group is then exposed to the change in the independent variable. A posttest measurement is then taken for both the groups.

Behaviors

Behaviors are the actions taken by respondents. Data related to consumer behavior is of great importance to marketers. Questions regarding respondents' behaviors toward a particular situation can be asked to them directly and can be included in a survey. However, the responses may not represent the actual behavior of the respondents. Observation techniques are more often used to understand the actual behavior of respondents. To further explore respondent behavior, marketers can categorize consumers according to product usage and user status.

Bibliographic Database

Bibliographic databases are generally composed of citations of journal articles, print media such as newspapers and magazines, and marketing research reports. Journals and periodicals are some of the most important sources for reliable data.

Binary Logistic Regression

This models the relationship between predictor(s) and a response that is a binary or dichotomous variable.

Binomial Distribution

A binomial distribution is a discrete probability distribution of the number of successes in a sequence of success/failure or yes/no experiments (e.g. tossing of fair coin n times under identical conditions).Conditions for the use of Binomial Distribution:

- Events are independent.
- Events are mutually exclusive with two outcomes (Success/Failure, Good/Defective, Yes/No, etc.).
- Probability (p) of each outcome should remain constant during trials.
- Number of trials is finite (n) for a particular experiment and n can take any non-negative integer value.

Bivariate Data Analysis

Bivariate analysis involves analysis of two variables or factors simultaneously. Because bivariate analysis involves two variables, the researcher is interested in identifying the relationship between them (i.e., whether or not one variable is dependent on the other and the nature of the relationship). Instead of measuring central tendency and dispersion or variance, as in the case of univariate analysis, the researcher will be measuring the degree of linear relationship between the variables (i.e., the correlation coefficient).

Branding and Advertising

Branding is the process of creating a distinct image of a product or range of products in the customer's mind. Advertising is defined in the SMstudy® Guide as any paid form of non-personal communications to existing and potential customers that promote the company's products through all media—such as radio, television, and print.

Business Unit Strategy

The Corporate Marketing Strategy, which is a component of the overall Corporate Strategy, is further divided into various Business Unit or Geographic Strategies, which in turn is further divided into particular Product or Brand Strategies for each product or brand.

Cartograms

Cartograms are maps that are used to present statistical data that is geographically based. They are maps in which geographic entities such as states or countries are resized according to a particular parameter of interest, such as population, sales, market share or electoral votes. Cartograms are very eye-catching. They are especially useful where geographic comparisons are significant and where approximate measures are acceptable.

Categorical Scales

Categorical Scales measure variables with a few distinct levels of measurement for values. For example the taste of a newly launched beverage can be measured as sweet, bitter, or bland. Two types of categorical scales are Nominal and Ordinal scales.

Causal research design

Causal research design is used when a cause-and-effect relationship has to be established between two variables related to the research problem. Put simply, when one independent variable (X) affects the state of another dependent variable (Y), there is a causal relationship.

Central Limit Theorem

The Central Limit Theorem states that when plotting the mean values of the samples taken from any population (which does not need to be normal), then the distribution of the sample means tends to a normal distribution, as the sample size increases.

Central Location Test

Central Location Test is a quantitative technique in which face-to-face interviews are conducted at a centralized location such as a room or booth in a shopping mall. It is considered an effective method for testing concepts, products, packaging, and advertising effectiveness, as well as conducting sensory research.

Chi-squared Test

A chi-squared test is applicable when data are categorical or when data are obtained by count (i.e., data consist of frequencies).

Chi-squared Distribution

A Chi-squared distribution is a right-skewed distribution: the curve has its long tail toward the higher values of the distribution. Some characteristics of this distribution include the following:

- Tests hypotheses regarding population variance against known or assumed value
- Assesses two types of comparison:
    - Chi square test for goodness of fit of an observed distribution to a theoretical one
    - Test of independence, which assesses whether paired observations on two variables, expressed in a contingency table, are independent of each other
- Has one parameter k, which is a positive integer that specifies the number of degrees of freedom

Choice or Ordering Techniques

This is a type of projective technique which requires the respondent to select from a set of alternative arrangements that fit some specific criteria, such as relevance, attractiveness or meaningfulness (e.g., a picture arrangement test).

Cluster Analysis

Cluster analysis is used to create groups of objects having similar characteristics. A large number of objects are divided into some groups. These groups are known as clusters. These clusters have certain properties. Within a cluster, the elements or objects are homogeneous in nature, and are heterogeneous among objects belonging to different clusters. The clustering process (i.e., grouping objects into different clusters) depends on the similarity of characteristics among the objects.

Cluster Sampling

In this sampling technique, the total population is divided into groups or clusters. These individual clusters are essentially similar to each other. A random sample of these clusters is selected.

Complementary Events

An event A is said to be a complement of event B if one and only one of them must occur, but they cannot occur simultaneously. All complementary events are mutually exclusive. When these two criteria apply, A and B are said to be complementary events.

Complementary opposition

In a 2X2 matrix that uses complementary opposition, qualitatively different options that can be interdependent are explored.

Completion Technique

This is a type of projective technique where subjects are provided with an incomplete stimulus situation and asked to complete it in any way they wish with the constraint that the completions meet certain standards of rationality or form. For example, a story completion may require a chronologic order

Computer Processing

Computer processing uses computer software to summarize, analyze, and convert data into usable information. Computers help in terms of accuracy, efficiency, and sophistication. For large amounts of data or surveys, computer processing becomes both convenient and economical.

Computer–assisted Personal Interview (CAPI)

The Computer-assisted Personal Interview (CAPI) is a quantitative technique in which respondents provide their answers on a computer in the presence of an interviewer. It is considered a personal interviewing technique because the interviewer accompanies the respondent during the process.

Computer-assisted Telephonic Interview (CATI)

The Computer-assisted Telephonic Interview (CATI) is a quantitative technique that uses computer software and a telephone system to conduct the questionnaire and record the responses. The computerized questionnaire is administered by the interviewer to the respondents over the phone. The interviewer records the responses directly into the software. This technique enables data analysis without the need for transferring responses to the computer and makes it possible for researchers to generate interim reports.

Computer-assisted Web Interview (CAWI)

The Computer-assisted Web Interview (CAWI) is one of the most commonly used quantitative techniques in recent times. The questionnaires are created in software specifically built for web interviews. This interface allows for the insertion of audio-video and links to different web pages. The program changes the flow of questions depending on the respondent's answers and the available information about the respondent.

Concomitant Variation

It is the extent to which the two variables in a cause-effect relationship vary together in a predictable manner.

Conditional Probability

The conditional probability of an event A, given that another event B has already occurred, is defined by $P(A/B)=(P(A \cap B))/(P(B))$

Conjoint Analysis

Conjoint analysis is a multivariate technique that quantitatively analyzes the joint effect of multiple attributes or features of a product or service that a customer desires.

Construction Technique

In this type of projective technique, respondents are required to go beyond simple association and to construct or create a more elaborate product, which is usually a complete art form, such as a story, description, dialogue, or picture. In a construction technique, the initial structure provided to the respondents is less than is the case with a completion technique. For example, the researcher might ask respondents to create a collage of what they feel when they experience a product, service, or concept under study

Contingency table

A contingency table is a two-way table of counts obtained when sample items are classified using two categorical variables. Each variable may have two or more categories. It is a test to verify whether or not the two classification criteria are independent.

Continuous data

Data generated by using some measurement device or physical instrument is known as continuous data. Continuous Data:

- Is measured on a physical instrument
- Can be a fraction of a whole number
- Can be further broken down into smaller parts

Examples: length, height, width, volume, time, temperature, viscosity, and velocity

Continuous Panels

In a study that uses a continuous panel, the questions asked of the panel members remain the same over the research period. This is helpful in problems where the researcher wants to understand the changes in customer behavior over a period of time.

Continuous random variable

If the set of possible values of a random variable is an interval or a collection of non-overlapping intervals, then the random variable is said to be a continuous random variable. An interval is defined as a set of all values ranging from the lowest number to the highest number.

Continuous rating scale

The continuous rating scale, also known as the graphic rating scale, is used where the respondents are asked to mark a point on a scale that they find appropriate. They are used to measure categories such as satisfaction, brand awareness, and quality in relation to a particular product or service.

Convenience sampling

Convenience sampling is widely used in student research projects. In convenience sampling, elements that are easy to get are chosen from the population for the study. Thus, this strategy may lead to biased data.

Conventional Mass Media Marketing

Unlike a seller's marketplace where sellers have the advantage over customers, mass media marketing features multiple manufacturers, thus shifting the balance of power in favor of consumers. Primary channels used for mass media marketing are print advertising (newspaper, magazine, insert, or run of paper), mass mailers, television (network, cable, or syndicated), and radio (national, local, satellite, or podcast), and outdoor advertising (billboards, bus shelters, stadiums).

Corporate Finance Strategy

This defines how the company will manage its finances, attain funding, and financially sustain its operations. The Finance Strategy should include forecasts and projections and summarize costs, income, and investments.

Corporate Human Resource Strategy

This maps the human resource capabilities within the company and considers talent management and acquisition needs to sustain growth.

Corporate Marketing Strategy

This defines how the company will target, position, market, and sell the planned products, and defines metrics, targets, and budgets for all marketing activities.

Corporate Operations Strategy

This defines how the company will manage operational activities, manufacture its products (or provide services), and provide the corresponding customer support and warranty.

Corporate Product Strategy

This defines the products or services the company offers, and the research and development efforts required to create them.

Corporate Strategy

Corporate Strategy is the overall direction of the company (as defined by senior management) that takes into consideration an assessment of the existing capabilities of the company and external opportunities and threats.

Correlation

Correlation examines the relationship between two variables using a correlation coefficient measure that indicates the strength of a relationship.

Correlation coefficient

A correlation coefficient is a measure of strength of a linear relationship between two variables denoted by r with the given range $-1 \leq r \leq 1$.

Critical region

The set of values of the test statistic for which the null hypothesis is rejected.

Cross-sectional Design

In cross-sectional design, the collection of information from the sample group happens at one point in time. Research to understand the customer preference for a particular brand while purchasing garments during a festive season would fall under cross-sectional design.

Data

Data are pieces of information pertaining to a particular subject(s) of study. Data are collected by taking measurements of selected attributes or features of the subject(s) under study. Data can be numerical or non-numerical.

Data cleaning

Data cleaning is the process of checking the data for omissions, consistency, and legibility. This step can involve, for example, checking for errors and omissions on questionnaires or other data collection forms.

Data coding

Data coding is the process of organizing, labeling, and sorting the data into appropriate headings to make it convenient for data analysis in both quantitative and qualitative research. In data coding, responses are usually assigned numbers and symbols to make them measurable and recordable.

Data Mining

Data mining is the most used secondary data processing tool. Through data mining, researchers can discover unknown valid information from huge databases; these discoveries help organizations make key business decisions. In other words, data mining is an exploratory data analysis without any prior hypothesis. Data mining is also an automated process of data analysis where the software or the system fetches valid information from the database.

Database Segmentation

It is used to classify the data into clusters. The segmentation is done using statistical cluster analysis techniques.

Degrees of Freedom

It is defined as an unrestricted variable (i.e., free to vary) in a sampling distribution. This variable can take any positive integer value.

Demographic Data

Demographic data are related to characteristics such as the gender, age, income, education, occupation, marital status, ethnicity, and social status of the target group. Demographic data are important because it helps marketers profile a target group. Demographic primary data help group the respondents into consumer segments.

Dependent Events

Two events A and B are said to be dependent if the occurrence of A depends on the occurrence of B and vice versa.

Dependent variable

The dependent variable is the outcome of the experiment that is measured by the researchers. Sales, awareness, and profit are typical dependent variables.

Descriptive Research Design

Descriptive Research Design is used to determine specific characteristics of a group, people, or organization associated with the marketing research problem. Descriptive research is most suitable if the researcher wants to know who the customers are, where they are from, and what they want. It provides answers to the questions who, what, when, where, and how. In this research design, basic information about the problem is available and the scope of the problem is limited to a few areas that need to be examined.

Deviation Detection

This method helps researchers determine the data or records that cannot be considered or included in the analysis. In this method data are identified and removed if they are not within the scope of the research problem.

Digital Marketing

It includes all marketing activities that use electronic devices connected to the internet to engage with customers (e.g., computers, tablets, smartphones). These include activities related to creating and managing effective websites and mobile apps as well as promoting a company's products and brand through various online channels that help meet marketing objectives.

Direct opposition

In a 2X2 matrix that uses direct opposition, the team examines opposing options from which to choose—i.e., selecting one option from options that are distinct from one another.

Discontinuous Panels

In a study that uses a discontinuous panel, panel members respond to different questions at different points in time over the research process. Since the panel members are available for the extended duration of the research, discontinuous panels act as a ready source of information for the researchers to address a wide variety of problems. In discontinuous panels, the demographics and characteristics of the panel members are representative of the target population. This ensures the accuracy of the results when applied to the larger group.

Discrete Data

Data generated by counting is known as discrete data. Discrete data:

- Can be produced by counts
- Can be classified/categorized into certain groups

Examples: number of visitors to a website, number of people who visit a retail storefront, number of inbound calls, number of times customers repurchase products, and binary classifications (e.g., good/bad, yes/no, pass/fail)

Discrete random variable

If the set of possible values of a random variable is a countable set, then the random variable is said to be a discrete random variable. A countable set is a set that can be counted. A countable set may be finite or infinite.

Discriminant Analysis

Discriminant analysis is similar to multivariate regression analysis in terms of usage. Both of these techniques are useful in forecasting or predictive modeling. The only difference between them is that the dependent variable is continuous in regression analysis, but for discriminant analysis the dependent variable is a categorical variable. Market researchers use discriminant analysis when there is only one dependent variable that is categorical and two or more independent variables. The dependent variable can have two or more categories.

Distribution Strategy

The Distribution Strategy for a product or service identifies the key areas involved in delivering the product or service to customers.

Elimination of Extraneous Factors

To study the effect of an independent variable on the dependent variable, it is imperative that the extraneous variables are controlled. A causal relationship can only be established if all the other variables except the ones under study are kept the same.

Essential Marketing Aspects

Marketing Strategy and Marketing Research are referred to as Essential Marketing Aspects. Both of these Aspects are mandatory and should be used to define, measure, and provide direction for the overall marketing efforts of a company.

Estimation

In real-life problems, the true values of each of a population's parameters (e.g., population mean and variance) are not known, and statisticians usually determine an estimated or approximated value using samples drawn from the population. In the process of determining an estimated value of a population parameter, an error is expected to arise. The theory of estimation is all about minimizing the error and establishing a measure of accuracy when estimating population parameters.

Ethnographic research

In ethnographic research, a researcher observes or interacts with a particular geography or ethnicity to determine consumer behavior. Understanding consumer rituals is essential in determining the marketing plans for a particular geography. Ethnographic research can help companies gain popularity and increase sales by understanding consumer rituals and preferences.

Event Space

An event space is a subset of the sample space. It is the set corresponding to the event under consideration. It is denoted as E.

Experiment

The alteration or manipulation of an independent variable to measure the impact on a dependent variable in a controlled setting constitutes an experiment.

Experimental design

Experimental design is one of the primary methods of causal research. In experimental design, independent variables can be manipulated in a controlled environment. This ensures that the effect of other variables on the dependent variable is minimized as much as possible.

Expert survey

This survey involves collecting data from a group having experience with the particular research problem.

Exploratory Multivariate Analysis

This is a quantitative technique that involves analyzing more than two variables or factors.

Exploratory Research Design

Exploratory research design is conducted for a research problem when the researcher has no past data or only a few studies for reference. Sometimes this research is informal and unstructured. It serves as a tool for initial research that provides a hypothetical or theoretical idea of the research problem. It will not offer concrete solutions for the research problem. This research is conducted in order to determine the nature of the problem and helps the researcher to develop a better understanding of the problem.

Exponential Distribution

Exponential distribution describes the time between events in a Poisson process (i.e., a process in which events occur continuously and independently at a constant average rate). It has the key property of being memoryless. It is used to model lifetimes of certain elements, to determine the average time between failures or the average time between a number of occurrences, and to analyze reliability.

Expressive Technique

In expressive techniques, respondents are given a visual or verbal situation and asked to relate the attitudes and feelings of other individuals to the situation (e.g., a role play). This is a type of projective technique.

External Validity

External validity refers to the extent to which the results of the experiment apply to the real world. This depends mainly on how well the test unit represents the actual population.

Extraneous variables

Any variable other than the intended independent variable that can affect the behavior of the dependent variable is called an extraneous variable. Extraneous variables need to be controlled in an experiment to study the effect of the independent variable on the dependent variable.

F Distribution

F distribution is used to test a hypothesis of equality of two population variances. Other characteristics of this distribution include the following:

- It is important in ANOVA, which is a technique used in design of experiments (DOE) for testing significant differences in variance within and between test runs
- F distribution has two parameters n1 and n2 called degrees of freedom
- The shape of the distribution curve is non-symmetrical and varies with the degrees of freedom

Factor Analysis

Factor analysis is a technique where all the variables are treated equally (i.e., the concept of dependent or independent variables does not exist). It is an interdependence multivariate technique. The main objective of factor analysis is to reduce a large number of variables into a relatively small number of factors by considering interdependence among all the variables simultaneously.

Fishbone

A Fishbone diagram, also called a cause and effect diagram, is a team tool for identifying possible causes to a problem and is often used in quality assurance measurement.

Focus groups

These are groups recruited to discuss the current research problem. They can provide insightful information and possible suggestions to support the marketing research process.

Fragmented New-Age Marketing

In recent times, the media has become increasingly fragmented with several hundred television and radio channels, as well as a large variety of print media, including newspapers, magazines, and trade publications. Moreover, since the late 1990s, with the increasing popularity of the Internet and, more recently, smartphones, many options now exist for advertisers to reach a global audience using digital marketing channels such as website, mobile, and social media. With all of these options, many marketers find it beneficial to use an integrated approach to marketing by leveraging the strengths of various types of media.

General media

General media such as the Internet, online databases, journals, broadcasts, and print media are major sources of external secondary data. A researcher can access a variety of websites, journals, and online versions of traditional print media to obtain required data.

Geographic Strategy

The Corporate Marketing Strategy (which is a component of the overall Corporate Strategy) is further divided into Business Unit or Geographic Strategy, which in turn is further divided into Product or Brand Strategy for each product or brand.

Government Publications

Government publications or reports may include statistics on production, agriculture, import, export, population census, and expenditure surveys. Federal, state, and local government publications or agencies can also be useful sources of information.

Hypergeometric Distribution

Hypergeometric distribution is similar to binomial distribution but with finite population and without replacement. Let the objects be selected at random, one after another without replacement from a finite population of size N consisting of k successes and N - k failures. Here, the trials are not independent since the selection is made without replacement.

Hypothesis Testing

A hypothesis is defined as an assumption about a population parameter. Examples of parameters are population mean and proportion. Hypothesis testing involves the careful construction of two statements: the null hypothesis and the alternative hypothesis.

Independent Events

Two events A and B are said to be independent if the occurrence of A is not dependent on the occurrence of B and vice versa. If A and B are two independent events, then $P(A \cap B) = P(A)P(B)$ (i.e., the probability of the occurrence of both the events is obtained by multiplying the individual probabilities.)

Independent variable

An independent variable is the trigger that leads to the effect. Also called treatments, these variables are manipulated as part of the experiment, and the effect on the dependent variable is measured. Advertising expenses, discounts, and price levels are typical independent variables.

In-depth Interviews

These interviews are often unstructured and conducted by the researcher to gain expert insights or to gain consumer opinions through door-to-door interviewing. Although these interviews are unstructured, the questions and probing are focused on the research problem. The main focus of in-depth interviews is to seek answers and possible solutions for the research problem.

Intentions

Intentions refer to the anticipated future behaviors of an individual. This is a subject of interest to marketers who want to solve a research problem related to future consumption rate or demand. Although there will always be some differences between consumers' intentions and actual practices, sellers believe that expected future behavior is a useful indicator to evaluate several possible alternative offerings.

Internal Reports

Different functions or departments within an organization have valuable sources of past data. These data or reports are generally identified in the Management Information System (MIS). Examples of internal secondary data are financial reports, sales reports, past research studies, inventory, and miscellaneous reports. Internal secondary data is not limited to the above examples and can vary depending on the industry or the research problem.

Internal Validity

Internal validity refers to the extent to which the manipulation of the dependent variable can be attributed to the independent variable. As discussed earlier, extraneous variables can distort the results of an experiment by affecting the dependent variables themselves. It is important that the effect of such variables is controlled in order to study causality.

Interval Estimation

When an interval or range is determined within which the true value of the population parameter lies, there is a measure of accuracy or level of confidence that the estimated value of the population parameter falls within the interval. This interval is known as the confidence interval (CI) for the associated population parameter. The level of confidence is expressed as a percentage.

Interval Scale

An interval scale not only contains the characteristics of both the nominal and ordinal scale but also can be used to determine the distance between the items. The degree of difference can be expressed using this scale, but not the ratio. Temperature scales and dates are examples of interval scales.

Kurtosis

Kurtosis is a measure of the peaked-ness or flat-ness of the data distribution. If a large portion of the data is concentrated near the mean and very few elements are away from the mean, then the variance will be very small. In this case, kurtosis will be very high and the distribution graph will have a high peak. On the other hand, if most of the elements of the data set are far from the mean (i.e., concentration near the mean is less), then the distribution graph will be flat.

Latent difficulties

These are the not-so-obvious challenges faced by the organization. These become evident in due time if not detected and addressed promptly. Such difficulties can be identified and resolved by marketing research, especially exploratory research.

Levels of Sales and Marketing Strategy

The Corporate Marketing Strategy, which is a component of the overall Corporate Strategy, is further divided into various Business Unit or Geographic Strategies, which in turn is further divided into particular Product or Brand Strategies for each product or brand.

Likert scale

A Likert item is a statement that a respondent is asked to evaluate. A Likert scale is the sum of responses to several Likert items. The Likert scale is useful for evaluating a respondent's purchasing behavior, opinion of a product, or level of product satisfaction.

Linear Model

A linear model is defined as a relation between two variables where changes in one variable produce a proportionate change in the other variable. Mathematically, a linear model is expressed as $Y = a + bX$, where a and b are constants that need to be estimated from the data. The strength of the linear relationship between two variables is detected with the use of a scatter plot and correlation coefficient. After that, the linear model to be used in forecasting is determined.

Link Analysis

This approach involves determining the associations between data and sequential patterns. This method helps discover the association between two or more data records, which in turn helps determine a possible solution to the research problem.

Lognormal Distribution

A lognormal distribution is a right-skewed distribution with most of the data residing in the left tail. For this distribution, natural logarithms of the original data follow a normal distribution.

Longitudinal Design

In longitudinal design, the same sample group is examined and the information is collected over an extended period of research. This is helpful for researchers who want to track changes in the sample group over a period of time.

Management decision problem

The management decision problem or management problem states the business-related issues. For example, a sales manager wanting to increase market share for product X is a management problem.

Market Segments

The output of using any of the segmentation tools is a description of the various market segments a company wants to consider. The descriptions should contain the characteristics of each segment that differentiate one segment from another.

Marketing research problem

The marketing research problem states the information that is required to solve the problem. One of the research problems for this management problem could be "how can we encourage the sales team to generate more sales?"

Mean

The mean, also known as the arithmetic mean or average, is defined as the sum of all the members in a set divided by the total number of members.

Measurement

Measurement is the act of assessing the level of presence of certain characteristics in the variable under consideration. The characteristics measured are attributes or qualities of the variable.

Measures of central tendency

Measures of central tendency such as mean, median, and mode give an indication of the center/location of the data.

Measures of dispersion

Measures of dispersion such as range and standard deviation give an indication of the spread of the data from the central location.

Measures of symmetry

Measures of symmetry such as skewness and kurtosis give an indication of the shape of the distribution of data, i.e. whether the data is distributed symmetrically (evenly on either side of the central location) or whether a greater percentage of data is concentrated either to the right or left of the central location.

Median

The median is the middle value of a set whose values are arranged in ascending order. It divides the set into two equal halves. Values in one half are less than or equal to the median, and values in the other half are greater than or equal to the median.

Method of Least Squares

It is a method to determine the best fit line for a set of data points. It states that among all the possible lines that can be drawn for a set of points, the best fitted line is the one for which the sum of the squares of the distances of the points from the line is the minimum.

Mode

The mode of a set is the element with the highest frequency of occurrence. The mode is determined by the number of times an element repeats itself in a set. A set may have more than one mode if several values have the same frequency of occurrence.

Motivations

A person's actions are the reflection of his or her inner state. Marketers often want to know the motives that direct specific consumer behavior. When a respondent is asked about the factors that are likely to influence his or her decision, a researcher is trying to identify the motives that influence behavior. Motivations can include users' category, brand-purchasing motives, value systems, and perceptions among others.

Multicolinearity

Multicolinearity is defined as the presence of a high degree of correlation between several independent variables in a multiple regression model. It can produce unexpected results; for example, the model may fit the data well (high value of F statistic) even though none of the predictor variables are statistically significant in predicting Y.

Multiple Cross-sectional Design

When two or more groups are used from the target group to collect information, this type of research is referred to as multiple cross-sectional design. The collection of information happens only once in each sample. Sample surveys, usually conducted by various television news channels, are multiple cross-sectional designs where the sample groups are representative of the larger target group.

Multivariate Data Analysis

The analysis involving more than two variables or factors is called Multivariate Analysis.

Multivariate dependence techniques

The most favorable scenario in multivariate data analysis is to have one dependent variable and multiple independent variables (i.e., one particular factor that is influenced by multiple other factors). Another scenario could have two or more variables that are dependent on multiple other independent variables. These two scenarios require multivariate dependence techniques. Based on the number of dependent variables and type of data (categorical or metric), there are several multivariate dependence techniques, including the following:

- Multiple Linear Regression
- Discriminant Analysis

Multivariate interdependence techniques

In multivariate interdependence techniques, there exist no dependent or independent variables. Here, all of the variables are treated equally to discover the interdependent relationship. Among the multivariate interdependence techniques are the following:

- Factor Analysis
- Cluster Analysis
- Conjoint Analysis

Mutually Exclusive Events

Two events are said to be mutually exclusive if they can never occur simultaneously. Mathematically, two events, A and B, are mutually exclusive if $A \cap B = \{ \}$.

Nominal Logistic Regression

This models the relationship between predictor(s) and a response that has three or more outcomes that do not have an order (i.e., pass, fail, defer).

Nominal Scale

A nominal scale is used for labeling variables. It is the simplest form of scale and can only be used to carry out the simplest of operations, such as categorizing variables into subsets.

Nonlinear Regression

This models the relationship between predictor(s) and response when quadratic or cubic terms are not adequate. It is used for nonlinear relationships, such as nonlinear growth or decay.

Non-probability Sampling Strategies

Non-probability sampling strategies are not as reliable as probability sampling strategies. The selection procedures in these strategies involve non-random methods. As a result, the subjects in the population do not have an equal chance of being selected as part of a sample. These types of sampling strategies are less likely to produce representative samples than probability sampling strategies. Regardless of this factor, many researchers have successfully used and continue to use these strategies.

Non-respondent Error

Also known as fieldworker error, these errors are made by the fieldworkers or administrators who administer the primary data collection by conducting interviews, surveys, or focus group discussions. Even with the use of professional data collection personnel, it is impossible to completely avoid this type of error. The errors may be intentional or unintentional. Such errors might include the incorrect recording of a subject's demographic information or a typographical error during data entry.

Normal Distribution

The graph of a normal distribution looks like a bell curve that is symmetric about the mean μ .Properties of a normal curve include the following:

- It is bell shaped (i.e., most of the data points are concentrated near the mean).
- It is symmetric about the mean.
- Mean = median = mode.
- The area under the normal curve = 1.
- It is unimodal (i.e., has one mode).

Null Hypothesis

This is the hypothesis about the population that needs to be verified. It is always about a population parameter and not about a sample statistic. The null hypothesis is denoted by Ho.

Observation Research

Observation Research involves the collection of information with regard to the behavior of individuals, objects, and organizations without any questions being asked of the participants.

Obtrusive Observation

These are commonly used within a sample group to understand consumer reaction to a product, commonly referred to as obtrusive technique. In this scenario, the consumer or the sample is aware of the observation.

One-way ANOVA

ANOVA can be applied in a one-way or two-way classification. The one-way classification treats data with only one classification criterion.

Online Marketplaces

Several e-commerce companies have created global online marketplaces for selling books, consumer goods, and other products. In such business models, customer acquisition is usually initiated through the company's website. The company coordinates with its multiple suppliers to source products, demos and product reviews are provided on the website, customers make their purchases online, and items are shipped directly to customers.

Online Networking

The Internet has made the world a smaller place. People can now have access to their networks at all times. These changes have significantly impacted the way in which people communicate with each other and, in turn, have created new possibilities for innovative business models.

Oral Report

An oral report is the presentation of research findings through verbal communication. The conclusions and recommendations of most research reports are presented orally as well as in writing. The purpose of an oral report is to communicate a project's the most important and provide stakeholders an opportunity to ask questions.

Ordinal Logistic Regression

This models the relationship between predictor(s) and a response that has three or more outcomes with an order (i.e., high, medium, and low).

Ordinal Scale

An ordinal scale is used for ranking the test units or their responses to the variable under consideration. It is used for ordering only and not for comparing.

Ordinary Linear Regression

This is the most basic type of regression between a predictor(s) and response variable. It models the linear relationship between predictor(s) and response when all data is metric and the response variable is continuous.

Orthogonal Regression

This models the relationship between one predictor and one response when the measurements of both the response and predictor include random error.

Outcome

The result obtained from a single trial of an experiment is called an outcome.

Overt difficulties

These are the challenges faced by the organization that are evident and need to be addressed immediately. For example, the reduced number of downloads for an e-commerce company's app is an overt difficulty. Such difficulties present the most urgent cases for marketing research.

Paper and Pencil Interview (PAPI)

The Paper and Pencil Interview (PAPI) is a quantitative technique that involves a direct interview of each respondent using a paper-based questionnaire. The use of PAPI has reduced greatly since the introduction of computers. However, it can still be convenient for non-complex questionnaires administered to small sets of people.

Pictograms

Pictograms are graphical displays used to present data. Pictograms are attractive and easy to understand, making them an effective method of communication. In pictograms, a pictorial symbol is used to clearly indicate the item that is being displayed.

Pilot survey

In a pilot survey, data are collected from a smaller sample group relative to the planned sample group. A pilot survey is a good option to test a survey before expense and time is invested in a larger more comprehensive survey.

Point Estimation

When a random sample is drawn from a population, population parameters can be estimated using sample statistics. The estimated value determined is the point estimate of the associated population parameter. The disadvantage here is that one does not know the level of accuracy for the estimate obtained.

Poisson Distribution

A Poisson distribution is the number of events that occur in a specified time interval with the following assumptions:

- Events occur singly (one after another) at a constant rate, λ per unit of time.
- The number of events that occur in non-overlapping time intervals is independent of one another.
- The events are described as occurring as a “Poisson process with rate λ.”

Poisson distribution describes the number of events that occur in a specified interval of time.

Porter’s Five Forces Model

Porter’s Five Forces model is used to analyze the long-term attractiveness of an industry. Understanding the interaction of these forces with the existing competing organizations helps explain the differences in profitability amongst industries. The five forces are Threat of New Entrants, Threat of Substitutes, Bargaining Power of Customers, Bargaining Power of Suppliers and Competitive Rivalry.

Posttest

The measurement that is taken after the exposure of test units to the independent variable is considered the posttest.

Power of a test (1 – β)

The probability of not failing to reject a null hypothesis when it is false.

Predictive Modeling

It is a form of inductive reasoning that takes specific data or information and makes a broader generalization. This technique uses neural networks and inductive reasoning algorithms to predict future trends. The conclusion from this method is always predictive and may not be accurate.

Presentation Software

There are various software packages that can be used in writing a research report or presenting it orally. There are many other online reporting programs that improve the efficiency of report writing and presentations to decision makers. There are several popular software tools that allow users to create different types of tables, charts, and graphics to present the content in a more effective manner.

Pretest

The measurement that is taken before the exposure of test units to the independent variable is referred to as the pretest.

Primary data

Primary data is newly obtained data for a specific purpose or a specific research project.

Probability

Probability can be defined as the possibility or likelihood of some event occurring. It is generally expressed in terms of a percentage, decimal, or fraction. The range of probability is from 0 to 1 (inclusive).

Probability Density Function (PDF)

The probability density function (PDF) is a continuous mathematical function, generally denoted as f(x) that models the probability density reflected in a histogram. PDF suggests the shape of the probability distribution.

Probability Sampling Strategies

Probability sampling strategies are the most reliable sampling strategies because the margin of error is minimal due to the statistical procedures used. In these strategies, every component in the population has an equal and independent opportunity to be chosen.

Process-Oriented Approach

In order to facilitate the best application of the SMstudy® Guide framework, the SMstudy® Guide defines a process-oriented approach to Sales and Marketing, which provides specific guidance to Sales and Marketing professionals about how to most effectively and efficiently manage their sales and marketing activities. The SMstudy® Guide defines Sales and Marketing in terms of processes that comprise a series of actions that leads to a particular result. Each process requires specific inputs and then uses tools to create specific outputs.

Product or Brand Marketing Strategy

Each Product or Brand Marketing Strategy (also referred to as 'Marketing Strategy' in the SMstudy® Guide) defines Sales and Marketing objectives for each product or brand, which drive specific tactics that align with and often rely on other Marketing Aspects (i.e., Marketing Research, Digital Marketing, Corporate Sales, Branding and Advertising, and Retail Marketing).

Projective Techniques

Projective techniques may be classified as a structured, indirect way of investigating the "whys" of a situation. Projective techniques are not used to measure, but to understand the attitudes, beliefs, and motivations of the consumer. Projective techniques help the researcher to understand customer perception of a product or a service.

Psychographics and Lifestyle Data

This kind of data is related to personality traits, interests, lifestyle, values, and opinions of the target respondents. Marketers often combine psychographics and lifestyle information with demographic information to obtain an important perspective of the target market.

Purposive sampling

A purposive sample is one that is selected based on the knowledge of a population and the purpose of the study. The main goal of purposive sampling is to focus on particular characteristics of a population. The subjects are selected because of the characteristic under consideration. Purposive sampling is applied only in situations where there are no other alternatives due to difficulties in locating and recruiting the desired population sample for the study.

P-value

In statistical hypothesis testing, the p-value is the probability of obtaining a result at least as extreme as the test statistic, given that the null hypothesis is true. Note that one rejects the null hypothesis if the p-value ≤ the significance level.

Qualitative Research

Qualitative Research uses unstructured approaches to obtain unquantifiable insights into behavior, motivations, and attitude from a small number of selected individuals.

Quantitative Research

Quantitative Research uses structured approaches to obtain measureable insights into the behavior, motivations, and attitudes of target respondents by involving a sample of the target population.

Quasi-Experimental Design

An experimental design that does not control the effects of the extraneous variables is called a quasi-experimental design.

Questionnaire

A questionnaire is a data collection tool that consists of a set of questions designed to obtain specific information from respondents. Correctly designing the questionnaire is a critical task because a properly designed questionnaire will enable the researcher to collect information that provides a decisive outcome. Questionnaires can be used to test hypotheses, gauge audience preferences, understand market demand for current product concepts, and decide on potential product features that may be attractive to the audience.

Quota sampling

Quota sampling is a non-probability sampling technique where the assembled sample has the same proportion of individuals as the entire population with respect to known characteristics, traits, or focused phenomenon. In quota sampling, a population is first segmented into mutually exclusive subgroups and then judgment is used to select the elements from a subgroup based on a specified proportion.

Random Variables

In statistics, a random variable is allowed to choose any value randomly within its permissible range. For example, a statistician is interested in studying the birth rate of a city and collects data on the “number of babies born per hour” for the past three months in that city. Let X denote the random variable “number of babies born per hour in the chosen city.” Therefore, X can take any non-negative integer value (i.e., possible values of X are 0, 1, 2, 3, 4, etc.).

Range

The range is calculated as the difference between the highest value and the lowest value of the set, also known as the extreme values. The range cannot be negative. It will be zero if there is only one value in the set, or if all the values are the same.

Rank Correlation Method

This method was developed by the British psychologist Charles Edward Spearman. This method is used for categorical data or ordinal data. There may be some factors for which quantitative measures are not possible (data is not continuous) but the data can be arranged in some order such that each individual is allocated a rank.

Ratio Scale

Unlike other scales, this scale deals with quantifiable data, and therefore has the concept of absolute zero. Not only does it contain the capabilities of the nominal, ordinal, and interval scales, but it also has the power of comparing the results using ratios. Actual measurements such as speed of a vehicle, age of a consumer, and number of children are all examples of ratio scales. Irrespective of the unit used to measure a particular quantity on a ratio scale, the ratio is always constant.

Reflexive opposition

In a 2X2 matrix that uses reflexive opposition, options in the same category are used on both axes (but they can be viewed from more than one perspective).

Region of acceptance

The set of values of the test statistic for which we fail to reject the null hypothesis.

Regression Analysis

Regression analysis examines the relationship between one dependent variable and one or more independent variables. When there is only one independent variable it is said to be a simple linear regression.

Research Design

The components of the research design consist of an objective/theoretical framework, analytical models, research questions, and hypotheses. The chosen research design is a set of guidelines or a blueprint that specifies the methods and procedures for obtaining and analyzing the required information. The research design will provide a framework or plan of action for the research project. This can be called the "Master Plan" or "Project Plan" for the entire research project.

Respondent Errors

These errors occur as a result of something the respondent does or does not do and may occur irrespective of data collection methods. Respondent errors might also be intentional or unintentional. Such errors may be the result of misinterpreting a question or incorrectly completing a form.

Retail Marketing

It includes all marketing activities related to persuading the end customer to purchase a company's products at a physical retail outlet or store, and efficiently managing the supply chain and distribution channels to improve the reach and sales for a company's products.

Sample Space

The set of all possible outcomes of a random experiment is said to be the sample space for that experiment. It is denoted as S.

Sampling

A population is the entire group of objects having characteristics of interest under study. The subset of a population that is chosen for the study is known as a "sample." Choosing the portion or subset of a population is known as "sampling." The chosen sample must represent all or most of the features of the population from which it is chosen.

Scale Transformation

In many instances, collected data must be compared with another data set. In order to make the data suitable for comparison, the scale needs to be transformed to ensure comparability with the scale values of the reference data set. The process of transforming the data to a meaningful, comparable format is known as scale transformation.

Scatter Diagram

A scatter diagram is a graphical representation of the relationship between two variables, and it provides both a visual and statistical means to test the strength of a relationship. A scatter diagram is also referred to as a scatter plot.

Secondary data

Secondary data is that which has already been collected for purposes other than the problem at hand.

Secondary Data Analysis

The process of searching and interpreting the existing information is known as secondary data analysis. An analysis of the secondary data is often referred as the core of exploratory research.

Seller's Marketplace

The Industrial Revolution in the eighteenth and nineteenth centuries marked a shift to mass production in factories (e.g., textile manufacturing). During this time, transportation infrastructure improved significantly with inventions such as the steam engine and more efficient ships. The banking system was further developed and the exchange of money became easier. Communication was also substantially improved through the development of the postal system and the use of telegraphs. Furthermore, goods were produced more efficiently and economically in factories and could be sold to a wider market. This created the seller's marketplace. The main objective of the seller's marketplace is to establish a supply chain to procure products and then establish a distribution channel to sell the products to a wide variety of customers, often referred to as "mass marketing." Emphasis on branding and advertising is minimal in a seller's marketplace.

Senior Management Direction and Insights

This is provided by senior management based on their experience and insights related to the business.

Seven-point semantic differential scale

The semantic differential scale is a scale used to discover respondents' attitudes toward a product or service. This scale asks respondents to rate a product, service, brand, or company based on a seven-point rating scale that has two bipolar adjectives at either end. It is designed to measure the connotative meaning of objects, events, and concepts.

Simple Random Sampling

This is the most basic type of probability sampling in which each individual is chosen randomly, so that each individual has an equal and unsystematic chance of being selected at any stage. Simple random sampling can be conducted in the following ways:

- Random numbers: The best way to conduct random sampling is by using random numbers generated by computer or using a table of random numbers.
- Lottery method: In this method, each element of the population is assigned a name or number. Each name/number is written on a paper slip and folded to look identical to the other slips. All the slips are then deposited into a box and mixed thoroughly. The required number of slips is drawn randomly from the box.

Single Cross-sectional Design

This approach refers to a research design in which only one sample group from a larger target group is used for collecting information. In order for the researcher to obtain accurate results, the sample group selected should be representative of the target group and share similar characteristics, such as social and educational status.

Situation analysis

A situation analysis involves examining the external environmental factors and internal organizational capabilities that impact how a company operates. Since companies operate in dynamic environments, understanding the changing landscape and current trends that are impacting the business helps marketing researchers identify the existing and potential problems.

Skewness

Skewness is a measure of asymmetry in a data distribution. If a data distribution shows a tendency to cluster around the higher values, then it is negatively skewed and, if it shows a tendency to cluster around the lower values, then it is positively skewed. Skewness is defined as the ratio of third central moment and standard deviation cubed.

SMBOK® Guide

A Guide to the SMstudy® Sales and Marketing Body of Knowledge (SMBOK® Guide), also referred to as the "SMstudy® Guide," is a series of books that provide guidelines for the Sales and Marketing of products and services. It offers a comprehensive framework that can be used to effectively manage Sales and Marketing efforts in any organization.

Standard Deviation

Standard deviation (SD) provides the measure of dispersion from the mean. It is the square root of the variance. Therefore, variance and standard deviation can be denoted as, $SD=\sqrt{V}$

Standard error (SE)

The standard deviation of the sampling distribution of means indicates the standard error (SE), which is the amount of error that can occur when a sample mean is used for estimating a population mean. Standard error is calculated as standard deviation divided by the square root of n where n is the sample size.

Standard Error = (standard deviation)/$\sqrt{n}$

Standard Normal Distribution

A normal distribution with mean μ and standard deviation σ can be made standard normal distribution using a linear transformation $Z = (X - \mu)/\sigma$. If the random variable X represents the normal distribution with mean μ and standard deviation σ, then the random variable Z will represent a standard normal distribution. The standard normal distribution has a mean of 0 and a standard deviation of 1. This is known as standardization of a normal distribution. It is required for further statistical analysis because it makes the calculations easy.

Statistical Inference

Statistical inference provides interpretations of and about a population from a sample. In practice, it is often impossible to analyze the entire data of a population due to constraints such as time, cost, and manpower; when this is the case, a sample is selected that is treated as representative of the population, and further statistical analysis is performed to interpret the population's characteristics.

Statistical Packages

These packages are specialized computer programs for statistical analysis. Some of the packages require a code or program to be written to execute a statistical analysis, whereas other advanced packages are made very user-friendly and do not require any coding.

Statistical test

A decision function that takes its values in the set of hypotheses.

Stratified Sampling

In this sampling technique, the entire population is divided into relatively homogeneous groups. These groups are known as strata. A random sample is chosen from each stratum. The size of the sample corresponds to the proportion of that stratum in the population as a whole. Alternatively, equal numbers of elements are drawn from each stratum and a weight is given to the results according to the stratum's proportion of the entire population.

Stratum Chart

A stratum chart is a two-dimensional graph with time along the horizontal axis and values of the items plotted along the vertical axis. Stratum charts are a variation of a line chart and have similar characteristics.

Strength of Association

Strength of association is a measure using the coefficient of determination (r2) similar to the case of simple linear regression. This value can be obtained easily as an output for regression analysis for many statistical tools.

Structured observation

Structured observation, also referred to as systematic observation, is an observational data collection technique in which the researcher collects the information directly without the mediation of respondents, interviewees, and so on. In this structured technique, the data is collected according to predefined rules.

Student's t Distribution

Student's t distribution or simply "t distribution" is commonly used to determine the confidence interval of the population mean when population variance is unknown. Some characteristics of this distribution are as follows:

- It is used to test a hypothesis when the means of sample populations are compared
- The shape of the t distribution approaches the standard normal distribution as the sample size (n) increases
- Typically at $n \geq 30$, t and normal distribution will be the same
- t distribution has one parameter n, which is a positive integer that specifies the number of degrees of freedom (degree of freedom is defined as an unrestricted variable in a distribution)

Surveys

Surveys can provide quality information relevant to the research problem. The surveys can be conducted, usually within a target group of past customers or people who have enough knowledge about the product or service.

SWOT analysis

SWOT analysis is an important method of conducting situation analysis. The internal capabilities are studied by identifying strengths and weaknesses while the external environment is studied by identifying opportunities and threats.

Symptomatic situation analysis

Symptomatic situation analysis can be used as a step in identifying a research problem. The three symptomatic situations are overt difficulties, latent difficulties, and unnoticed opportunities.

Systematic Sampling

In systematic sampling samples are selected according to some specified systematic rule, such as the selection of elements from the population from a random starting point and at uniform intervals.

Tabulation

Tabulation is one of the most frequently used tools in data analysis. It is used to organize the data into predefined categories and to count the number of observations from the data set that occurs in each of the predefined categories. When only one variable is involved in the analysis, it is called univariate tabulation. When two or more variables are involved, it is called bivariate or multivariate tabulation, respectively.

Target Segment

Once a company has identified all market segments, explored the competition, and then compiled the details of competitive products, it should then analyze the various segments and the strengths, weaknesses, opportunities, and threats faced by the company in order to identify the target segments in which the business would be most competitive. This process involves identifying the type of customers a company plans to target and the product categories under which it intends to create products.

Test of goodness of fit

This is a test to verify whether it is reasonable to regard a random sample as coming from a particular specified distribution (i.e., whether or not a particular model provides a good fit to the data). In other words, the marketing researcher is going to test if there is any significant difference between an observed frequency distribution and a hypothetical frequency distribution. The researcher is interested in determining whether the observed sample corresponds with a Normal, Poisson, or Binomial distribution, and how well the observed pattern fits the expected pattern.

Test units

Test units are individuals, groups, or entities that are exposed to the experiment to measure their response on the dependent variable.

Time Order of Occurrence

Time order of occurrence of the cause and effect is of critical importance for causality. The cause X should precede or occur simultaneously with the effect Y.

Traditional Marketplace

Traditional marketplaces are usually small markets where price negotiations and other decisions related to sales are made quickly—often by one or two persons.

Trial

Any particular performance of a random experiment is called a trial.

True Experimental Design

An experimental design that eliminates the effects of the extraneous variables is called a true experimental design. Control groups are used to achieve this effect.

Two-way ANOVA

In this classification, data are classified according to two criteria of factors. In addition, when the researcher is conducting two-way ANOVA, it can be done with or without replication.

Type I Error

Also known as an error of the first kind, an α error, or producer's risk, it is the probability of rejecting a null hypothesis when it is actually true.

Type II Error

Also known as an error of the second kind, a β error, or consumer's risk, it is the probability of failing to reject a null hypothesis when it is false.

Unnoticed opportunities

Many opportunities are not very evident and some effort is required to identify them. Marketing research is invaluable in identifying and leveraging such opportunities.

Unobtrusive Observation

When the consumer is not aware of the observation, the technique is referred to as unobtrusive observation.

Unstructured observation

In unstructured observation, the researcher enters the field with some general ideas of what might be salient, but not of what specifically will be observed. Therefore, observation is holistic, unstructured, and unfocused, with the investigator attempting to document as much as possible about the setting and its participants in order to discover themes of interest.

Use Cases

A use case is a methodology or list of steps that defines the interactions between stakeholders and a system. A use case is an effective tool for identifying a research problem. Use cases can be used to describe how a customer's needs are met (or not met) by the product or service offering.

Validity of an experiment

The validity of an experiment is the extent to which the experiment measures what it is supposed to measure. Experiments can be deemed valid if the following two conditions are satisfied:

- if it is established that the change in dependent variable has indeed occurred due to the independent variable and not due to other extraneous variables
- if the results of the experiment apply to the larger population of interest or to other populations of interest

Variable Respecification

The main purpose of respecification is to transform the data to create new variables, or to modify existing variables, and ensure that these variables are more consistent with the research objective.

Variance

Variance (V) is the measure of the spread of numbers from the mean in a set. It considers all the values of a set, instead of just the extreme values, and calculates how close the values are spread around the mean. It is calculated as the average of the squared deviations of each value from the mean.

Variance Inflation Factor (VIF)

Variance Inflation Factor (VIF) is a measure of multicolinearity in a multiple linear regression model. Most advanced statistical tools produce VIF as an output of the regression analysis. VIF quantifies how much variance of the estimated coefficient is inflated due to the presence of multicolinearity.

Weibull Distribution

The Weibull distribution is used in analyzing reliability. It can also be used in applications that are similar to those that use a lognormal distribution (e.g., to measure such values as time to fail, time to repair, and material strength).

Weighting

Weighting is a tool used to assign each respondent in the database a value or multiplier that accounts for the extent to which certain categories of respondents should contribute to the overall results. Weighting is mostly used to make the sample data more representative of a target population.

INDEX

2

5

A

B

C

D

E

F

G

H

I

J

K

L

M

N

The SMstudy® Guide series is developed by VMEdu, Inc., a global education and certification course provider that has educated over 400,000 students world-wide in more than 3,500 companies. It is the result of a collaborative effort from a large number of Sales and Marketing experts with extensive experience, knowledge, and insights from a variety of industries and disciplines.

This second book in the SMstudy® Guide series focuses on Marketing Research, providing a unique process-oriented framework that, when effectively used, ensures a thoughtful and methodical approach to planning and executing Marketing Research projects. These projects, if implemented successfully, provide critical insights for taking key marketing decisions.

Marketing Research is one book in a series of six. Others in the SMstudy® Guide series that can be used to complement the initiatives arid tools detailed in this book are Marketing Strategy (MS), Digital Marketlig (DM), Corporate Sales (CS), Branding and Advertising (BA), and Retail Marketing (RM).

SMstudy
Targeting success

SMstudy® is a brand of VMEdu, Inc.