ISBN: 978-81-971650-3-0

8. Processing and Analysis of Data

Dr. Gagan Gunjan

Asst. Prof.,
Medicine RIMS Ranchi,
Jharkhand.

Abstract:

Since the beginning, the research process and data analysis have been extensively studied in academic and business settings; yet, the majority of students are uninterested in learning the concept, necessity, and direction of the research from books. In order to progress to a new area of study in any aspect of life, students must become aware of the difficulties related to the decision-making process and learn how to identify, describe, and solve them.

This study aims to review the procedures involved in data analysis and the research process. This study provides guidance on choosing a research design, creating a conceptual framework, and organizing data analysis. This research highlights the importance of adhering to research ethics throughout the entire process. It also provides assistance to researchers and students in clarifying the problem, defining the goals of the study, identifying the variables and connecting them to the goals of the findings, and gathering and analyzing data to generate insightful recommendations from the research findings. We'll talk about in this paper. Data analysis and processing.

Keywords:

Processing, Analysis, Data, Decision Making Process, Editing, Coding, Collection, Preparation, Input, Output, Storage, Classification, Tabulation, Nominal, Cohort Analysis, Cluster Analysis.

8.1 Introduction:

The term "data processing" describes a number of procedures, including coding, editing, scoring computation, creating master charts, etc. Each and every step of the research process requires a strategy from the researcher. Because of this, a skilled researcher creates an ideal plan for data processing and analysis. Data processing and analysis may not seem like extremely serious tasks to some researchers. They frequently believe that computer assistance handle data processing.

8.2 Methods for Data Processing in Research:

The gathering and conversion of a data set into useful, usable information is known as data processing in research. By using this technique, a researcher, data engineer, or data scientist can take raw data and use an automated tool or manual labor to transform it into a more understandable format, like a graph, report, or chart.

The knowledge will then be used by the researcher to make adjustments, discover new perspectives, resolve issues, and eventually provide better outcomes.

8.2.1 Data Processing Cycle:

The cycle of data processing involves feeding raw data (input) into a system in order to extract useful insights (output). The process is carried out in a certain order for each stage, but it is repeated cyclically throughout. As the example below illustrates, the output of the first data processing cycle can be saved and used as the input for the subsequent cycle. [1]

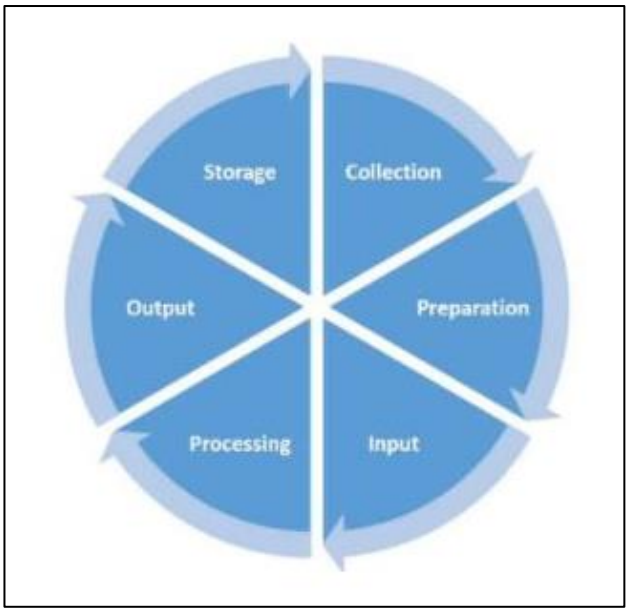

Figure 8.1: Data Processing Cycle

- **Step 1: Collection**

The practice of gathering data from accessible sources, such data lakes and warehouses, is known as data collecting. A variety of formats, including user behavior, financial data, profit statements, and online cookies, might contain raw data.

The kind of raw data you gather will greatly influence the final product that you create. For reliable, actionable research results, researchers need to consult reliable, thorough, and accurate sources.

- **Step 2: Preparation**

You will clean up, arrange, filter, and check raw data for faults through data preparation. The goal of the data preparation step is to remove inaccurate, unnecessary, or partial data and transform it into a format that is appropriate for additional processing and analysis. Obtaining the best quality data is the aim of the preparation phase.

- **Step 3: Input**

The input stage is when unprocessed data first starts to take on the characteristics of useful information. Once the data is clean, you'll translate it into a language that these systems can understand and enter it into the appropriate location, such a data warehouse or customer relationship management (CRM) program. There are other ways to enter this data, such as keyboards, scanners, and digitizers.

- **Step 4: Data Processing**

This step involves applying machine learning and artificial intelligence algorithms to the raw data in order to produce a desired result through a variety of data processing techniques. Depending on the data source (data lakes, online databases, linked devices, etc.) and the intended use of the output, this phase may differ slightly between processes. [2]

- **Step 5: Output**

Finally, the data is sent and presented to the user in a readable format, such as tables, graphs, audio, video, documents, and vector files. In the subsequent data processing cycle, this output might be kept and processed once more.

- **Step 6: Storage**

All information that remains after the data has been correctly processed should be saved for further use. Businesses that keep their data appropriately maintain compliance with data protection laws and facilitate quicker, simpler access to information when needed. This data can also be used as an input in the subsequent processing cycle.

8.2.2 Operations in Data Processing:

A. Editing:

a. The initial stage of data processing is data editing.
b. Examining the gathered raw data to find mistakes and omissions and fixing them when feasible is the process of editing data.
c. A thorough examination of the completed schedules and/or questionnaires is done during the editing phase.
d. Data editing is done to make sure the information is correct, consistent with other information, consistently entered, and potentially complete. It is also seen that the data have been organized neatly to aid in the subsequent stages, such as coding and tabulation, during the editing phase.
e. Thus, editing entails carefully examining the gathered data to find and, to the greatest extent feasible, reduce errors, incompleteness, misclassification, and gaps in the information received from respondents.
f. In large-scale surveys, editors or supervisors are appointed by the study project's sponsor to edit the data.

Editing is divided into two categories based on the stages at which it is completed: A. Editing in the field B. Editing Center. [3]

B. Coding:

The process of giving responses letters, numbers, or other symbols such that they may be classified into a restricted number of classes or categories is known as coding. These courses ought to be relevant to the current research issue. They must also have the properties of mutual exclusivity, which allows a particular response to be placed in one and only one cell within a given category set, and exhaustiveness, which requires that there be a class for each data item. One other criteria to follow is one-dimensionality, which states that each class should be defined in terms of a single idea. Coding is essential for effective analysis because it allows the many replies to be condensed into a small number of classes that hold the crucial data needed for analysis.

Usually, coding choices should be made during the questionnaire's design phase. Due to the ability to directly extract crucial information from the original questionnaires, this facilitates the preceding of the questionnaire options, which is beneficial for computer tabulation. However, if hand coding is required, a conventional technique may be applied. Using a colored pencil to code in the margin is one such common practice. The data from the questionnaire can also be transcribed to a coding sheet as an alternative technique. It is important to ensure that coding errors are either completely eradicated or drastically decreased, regardless of the approach used.

C. Classification:

In order to obtain significant associations, the vast amount of raw data generated by most research investigations must be condensed into homogeneous groups. This fact makes data categorization necessary, which is the process of grouping data into classes or groups based on shared features. Data that share a trait are grouped together into a class, which further divides the total data into multiple groups or classes.

Depending on the type of phenomenon involved, classification can take one of the two following forms:

Classification based on attributes: As mentioned previously, data are categorized based on shared traits, which can be numerical (e.g., weight, height, income) or descriptive (e.g., literacy, sex, honesty, etc.). Descriptive attributes are qualitative phenomena that are only observable in an individual item and cannot be quantitatively assessed. This type of data collection results in statistics of attributes, and the classification of such data is referred to as classification according to attributes. [4]

D. Tabulation:

Once a large amount of data has been gathered, the researcher must organize it in a clear, logical sequence. The process known as tabulation is this one. Therefore, tabulation is the process of condensing raw data and presenting it for additional analysis in a compact format

(i.e., statistical tables). An organized configuration of data in rows and columns is known as tabulation in a more general sense. The following justifies the need for tabulation.

a. It saves room and makes fewer explanatory and descriptive statements.
b. It makes the comparison process easier.
c. It makes item summaries and error and omission detection easier.
d. It gives different statistical computations a foundation.

Electronic, mechanical, or manual devices can all be used for tabulation. The size and type of the study, financial constraints, scheduling demands, and the accessibility of computers or tabulating equipment all influence the decision. If other conditions are met and the required resources are available, we may employ mechanical or computer tabulation in somewhat big questions. When conducting small-scale research with a limited number of surveys and short questionnaire lengths, hand tabulation is typically recommended.

The list and tally, card sort and count, and direct tally methods can all be used for manual tabulation. Tailoring the tally directly from the questionnaire is possible when there are simple codes. Using this method, the codes are written on a tally sheet, a piece of paper, and each response is marked with a stroke against the code that corresponds to it. Typically, a diagonal or horizontal line through the strokes indicates the fifth response, which comes after every four strokes against a specific code. The data is conveniently sorted against each code, and counting these groups of five is a breeze. By using the listing method, each questionnaire can have its own line when the code responses are transcribed onto a huge worksheet. In this manner, a work sheet can list a huge number of questionnaires.

Next, tallies for every question are created. The most adaptable way of hand tabulation is the card sorting technique. This method uses special cards with a number of holes that are conveniently sized and shaped for recording data. Every hole represents a code, and when the cards are stacked, a needle goes through the hole that corresponds to that code. After separation, these cards are counted. By repeating this method, one may thus determine the frequencies of different codes. If we need fast results, we can also use the computer facility or mechanical devices for tabulation; our budget allows us to use them, and we have a lot of simple tabulation with lots of cross-breaks. [5]

8.3 Data Analysis:

The process of gathering, modeling, and evaluating data using a range of logical and statistical procedures is known as data analysis. Companies use analytics tools and processes to gather data that helps them make operational and strategic decisions.

The two main pillars of all these different approaches are research, both qualitative and quantitative.

8.3.1 Quantitative Methods:

Quantifiable or numerical data can be interpreted using the quantitative data interpretation technique. There are two categories of numerical data:

- **Discrete:** Finite, countable amounts. For instance, the number of ice creams.
- **Continuous:** Not able to be tallied. For instance, temperature, humidity, time, speed, height, and weight.

The statistical modeling techniques, such as central tendency and dispersion measures, are comparatively easy to apply when analyzing numerical data. Charts like pie charts, bar graphs, line graphs, and line charts can be used to visually represent this. Tables are another useful tool for categorizing and displaying complex data.

A. There are two most commonly used quantitative data analysis methods are:

- **Descriptive Statistics:** The statistical field of statistics is concerned with characterizing the data and its attributes. It is divided into two categories: measures of dispersion or variability, which indicate how much the data fluctuates, and measures of central tendency, which include mean, median, and mode.
- **Inferential Statistics:** Utilizing a sample picked from the larger data, this area of statistics extrapolates or generalizes the features of the bigger data. [6]

8.3.2 Qualitative Methods:

To examine textual and descriptive data, also known as categorical data, qualitative approaches are used. Most text data are unstructured. On the basis of their attributes, the qualitative data is further divided:

- **Nominal:** There is no hierarchy or order among the qualities. For instance: Location, gender, and classes in school
- **Ordinal:** There is a ranking or ordering of the qualities. For instance: Grades
- **Binary:** There are just two types in it. Class 1 or Class 0, yes or no.

As the data in this case is non-statistical and the computers only speak the language of numbers, categorical data, unlike numerical data, cannot be immediately analyzed.

In order to transform the text data into numerical data, it must first be coded. Various coding techniques are available depending on the need. To be used in modeling and interpretation, the text data is labeled. [7]

8.3.3 Data Analysis Methods:

A. Regression Analysis:

A useful technique for figuring out how one or more independent variables and a dependent variable are related is regression analysis. It is used in the social sciences, finance, and economics.

Regression modeling allows you to identify patterns in your data, examine cause-and-effect relationships, and create predictions.

B. Statistical Analysis:

The field of statistical analysis comprises an extensive array of methods for condensing and analyzing data. It includes multivariate analysis, inferential statistics (confidence intervals, hypothesis testing), and descriptive statistics (mean, median, standard deviation).

Statistical techniques aid in drawing conclusions, evaluating the importance of findings, and drawing inferences about populations from sample data.

C. Cohort Analysis:

Understanding the behavior of particular groups or cohorts across time is the main goal of cohort analysis. Businesses can adjust their strategy by using the trends, retention rates, and client lifetime value that it can uncover. [8]

D. Content Analysis:

It is a method for qualitative data analysis that is used to examine the textual, visual, or multimedia content. It is frequently used in the social sciences, journalism, and marketing to examine patterns, themes, or emotions in texts or other media. Large amounts of unstructured data might provide valuable insights for academics when they use content analysis.

E. Factor Analysis:

Finding underlying latent components that account for the variance in observed variables is possible with the use of factor analysis. It is frequently used to uncover underlying constructs and reduce the dimensionality of data in psychology and the social sciences. Complex datasets can be simplified by factor analysis, which facilitates their interpretation and analysis.

F. Time Series Analysis:

Data that is gathered throughout time at regular intervals is the subject of time series analysis. Forecasting, trend analysis, and comprehending temporal patterns all depend on it. Autoregressive integrated moving average (ARIMA) models, moving averages, and exponential smoothing are examples of time series techniques. They are extensively utilized in economics for economic modeling, meteorology for weather forecasting, and finance for stock price prediction.

G. Prescriptive Analysis:

By making recommendations for decisions or actions based on the forecasts, prescriptive analysis goes beyond predictive analysis. To produce meaningful insights and maximize results, it integrates historical data, optimization algorithms, and business rules. It facilitates resource allocation and decision-making. [9]

8.3.4 Quantitative Methods:

In short, quantitative analysis encompasses any techniques that extract meaningful insights from numerical data or data that can be converted into numbers, such as category variables like gender, age, etc.

It is employed to draw insightful findings regarding similarities and differences as well as to test theories. We go over a few of the important quantitative techniques below.

- **Cluster Analysis:**

The term "cluster" refers to the process of organizing a collection of data items so that they are more comparable to one another (in a specific sense) than they are to those in other groupings. Clustering is a popular technique for uncovering hidden patterns in data because it doesn't have a target variable. The method can also be applied to provide a trend or dataset more context.

- **Cohort Analysis:**

This kind of data analysis technique compares and examines a certain user behavior segment using historical data so that it may be categorized with other users that share the same traits. This practice might help you grasp a larger target group or obtain a wealth of information into the needs of your customers.

- **Regression Analysis:**

Regression analysis makes use of past data to determine the effects of one or more independent variables changing or remaining constant (multiple regression) on the value of a dependent variable.

It is possible to predict potential outcomes and make wiser judgments in the future by comprehending the relationship between each variable and how it has evolved historically.

- **Neural Networks:**

The clever algorithms of machine learning are based on neural networks. This type of analytics makes an effort to comprehend how the human brain might produce insights and forecast values with the least amount of intervention. Neural networks are always evolving and improving because they take in knowledge from every data transaction.

- **Factor Analysis:**

A sort of data analysis known as "dimension reduction," or factor analysis, describes the variability among associated, observed variables in terms of a smaller number of possible unobserved variables known as factors. Finding independent latent variables is the goal here, as it is a perfect way to simplify certain parts.

- **Data Mining:**

An approach to data analysis that serves as a catch-all for engineering metrics and insights for extra context, value, and guidance. Data mining looks for relationships, patterns, trends, and dependencies in order to produce enhanced knowledge using exploratory statistical analysis. Adopting a data mining approach while thinking about data analysis is crucial to success; for this reason, it's a topic worth delving deeper into. [10]

8.3.5 Steps for Data Analysis Process:

- Data Requirements Specification
- Data Collection
- Data Processing
- Data Cleaning
- Data Analysis
- Communication

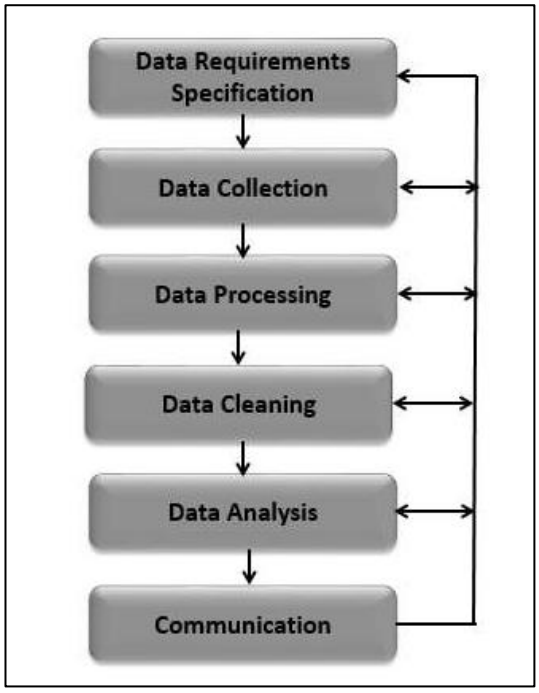

Figure 8.2: Steps for Data Analysis Process

8.4 Data Requirements Specification:

An experiment or a query serves as the basis for the data needed for analysis. The data required as analysis inputs are determined based on the requirements of the individuals guiding the analysis (e.g., Population of people). One can specify and retrieve certain population variables (such age and income). Data can be either categorized or numerical.

8.4.1 Collect Data:

The preparation or collection of the data is the second step. Data collection and storage for later analysis are part of this process. The analyst must gather information from several sources in accordance with the assignment assigned. It is necessary to gather the data from a variety of internal and external sources. The data that is available within the company you work for is known as internal data, while the data that is available outside of your company is known as external data. First-party data is information gathered by an individual using their own resources. [11]

8.4.2 Data Cleaning:

The Clean and Process Data stage is the third. Once information has been gathered from various sources, it's time to clean it up. Misspellings, redundancies, and irrelevant information are not present in clean data. Data integrity is a major factor in clean data. Unnecessary data is eliminated and cleaned since it may be redundant or not in the proper format.

8.4.3 Analyzing the Data:

Analyzing is the fourth step. For analysis and trend identification, the cleaned data is utilized. For improved outcomes, it also aggregates data and does calculations. SQL or Excel are the tools utilized to make calculations. These tools come with built-in functions for performing computations, or they can be used with sample SQL code. While SQL produces temporary tables for calculations, Excel allows us to build pivot tables and do calculations. Another method of problem solving is through programming languages. [12]

8.4.4 Data Visualization:

Visualizing the data is the fifth phase. There is nothing more powerful than a picture. Now that the data has been translated, a visual (chart, graph) must be created. Data visualizations are created because non-technical stakeholders may comprise some portion of the population. The purpose of visualizations is to make complex data easily understandable. The two widely used programs for creating eye-catching data visualizations are Tableau and Looker.

8.5 Communication:

In order to support the users' decisions and subsequent actions, the data analysis results must be delivered in a format that meets their needs. More research may be conducted as a result of user input.

Tables and charts are examples of data visualization tools that data analysts can employ to help users understand and process information more effectively. The ability to highlight necessary information in tables and charts using color coding and formatting is made possible by the analysis tools. [13]

8.6 Conclusion:

To sum up, the ability of data analysis procedures to condense complex information into understandable, visual narratives enables firms to make wise decisions. Effectively conveyed data-driven insights are essential for tackling business challenges and promoting ongoing growth in a variety of fields.

8.7 References:

1. Silverman, S., Manson, M. (2003). Research on teaching in physical education doctoral dissertations: a detailed investigation of focus, method, and analysis. Journal of Teaching in Physical Education, 22(3): 280-297.
2. Berg, B.L. (2001). Qualitative Research Methods for the Social Sciences. Boston: Allyn and Bacon.
3. Foster, A. (2004). A nonlinear model of information-seeking behavior. Journal of the American Society for Information Science & Technology, 55(3), 228-237.
4. Smeeton, N., Goda, D. (2003). Conducting and presenting social work research: some basic statistical considerations. Br J Soc Work, 33: 567-573.
5. Snider, J. (2010, February 10). Commentary. The cult of statistical pyrotechnics. Education Week, 29(21), 20–21.
6. Thompson, B., Noferi, G. 2002. Statistical, practical, clinical: How many types of significance should be considered in counseling research? Journal of Counseling & Development, 80(4):64-71.
7. Ahuja Ram. Research Methods. Jaipur: Rawat Publications, 2007.
8. Bryman Alan. Social Research Methods. New York: Oxford University Press, 2004.
9. Elifson Kirk W, Runyon Richard P. and Haber Audrey. Fundamentals of Social Statistics. United States: Mc. Graw Hill, 1997.
10. Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). Introduction to research in education. Cengage Learning.
11. Awang, Z. (2012). Research methodology and data analysis second edition. UiTM Press.
12. Blumberg, B., Cooper, D., & Schindler, P. (2014). EBOOK: Business Research Methods. McGraw Hill.
13. Neuendorf, K.A. (2002). The Content Analysis Guidebook. Thousand Oaks, CA: Sage Publications.