Basics of Data Decision- making in Business

Lazar Pavić
University of Maribor, Faculty of Logistics, Celje, Slovenia
lazar.pavic@um.si

In this chapter, the fundamental concepts necessary for successful business decision-making based on data are presented. The basic phases of research and fundamental concepts essential for effective decision-making are introduced: statistical population, units of analysis, representative sample, variables, measurement scales, and hypotheses. The importance of collecting primary and secondary data in business decision-making, along with proper tabular and graphical representation of data, is then emphasized. The chapter is concluded with guidance on selecting appropriate statistical analysis to confirm or reject properly formulated research hypotheses.

DOI
https://doi.org/
10.18690/um.fl.3.2026.5

ISBN
978-961-299-099-2

Keywords:
business decision-making,
data,
data processing,
data visualization,
statistical analysis

University of Maribor Press

1 Introduction

In today's business world, where information is rapidly generated and changing, data-driven decision-making has become a key component of successful business. The fundamentals of data-driven business decision-making provide the foundation on which organizations build their strategies and plans. This approach enables organizations to transform the information they collect into actionable knowledge that helps them make better decisions.

One of the key benefits of data-driven decision-making is its speed. The ability of organizations to quickly analyze and interpret data enables agility in responding to changes in the environment. This is especially important in dynamic sectors such as logistics and supply chain management, where market conditions change rapidly. Another important benefit is better targeting of marketing campaigns. By analyzing data, organizations can accurately understand the needs of their customers, which enables targeted marketing and personalized approaches. This increases the effectiveness of marketing efforts and improves customer satisfaction. The optimization of business processes is another key benefit of data-driven decision-making. By recognizing the efficiency of business processes and identifying areas for improvement, organizations can increase operational efficiency and reduce costs. By analyzing quality data, organizations also gain insight into patterns and trends, which allows them to adapt their offerings and services. Tracking customer preferences, changing market trends, and adapting business models are concepts that become possible when decisions are supported by quality data.

In addition, data-driven decision-making enables improved risk management. Identifying potential problems through the analysis of reliable data allows organizations to take preventive measures and reduce the negative impacts of potential risks. Accordingly, data-driven business decision-making also enables innovation. By constantly tracking data, organizations can quickly identify opportunities for innovation and adaptation, leading to sustainable competitive advantage.

In this chapter, we will explore key aspects of data-driven decision-making and present basic concepts for developing guidelines for the successful implementation of data-driven decision-making in a business environment.

2 Phases of data exploration

Data exploration takes place in three separate but interconnected phases (Košmelj, 2007):

- collecting and organizing data;
- analyzing the collected data; and
- explaining or interpreting the results.

The data collection and editing phase is crucial for the reliability of the final results. The process includes data collection planning, which determines the type of data, data collection methods, and sample. This is followed by quality control of the data collected, which includes assessing the adequacy of the information collected and identifying and correcting any errors. Data editing includes coding, classification, normalization, and format conversion for optimal analysis. Data preparation for analysis includes filtering, selecting key variables, partitioning data, and creating new variables. Proper implementation of these steps ensures the purity, reliability, and readiness of the data for accurate statistical analysis, which in turn leads to high-quality research results.

The analysis phase of the collected data in research is crucial, as researchers use various statistical methods to understand, describe and interpret the data. In this phase, the selection of the appropriate statistical analysis is crucial. The most commonly used analyses include descriptive statistics, identifying statistically significant differences between different groups of subsamples, correlation and regression analysis, etc. Tabulating the results and visualizing them with charts are key in gaining in-depth insight and drawing conclusions based on statistical analyses. However, it is important for researchers to remain cautious and take into account potential limitations and biases of the research conducted.

In the interpretation phase, researchers closely link the results obtained to the objectives of the study. It is important to assess the statistical significance of the results, verify the practical relevance of the results, and consider any limitations of the study, including alternative explanations. Linking to existing knowledge allows the findings to be placed in a broader theoretical framework, and while this phase is important for scientific research, it is usually not carried out in data-driven business decision-making. Correct interpretation is key to understanding the significance of the results and their application to the context of the research.

The rest of the chapter will introduce the basic concepts related to data-driven decision-making. Understanding these concepts is crucial and forms the basis for properly dealing with data.

2.1 Mass phenomenon

For the purposes of data-driven decision-making, we focus on mass phenomena. These are phenomena that occur in large numbers in a business environment at a certain time and in a certain place. For each mass phenomenon, we can ask at least one research question, as explained below. In this way, we examine all the rules and laws that generally apply to the mass phenomenon, while at the same time learning very little about the specific individual case (Košmelj, 2007).

Example of a mass phenomenon: employees in the field of supply chain management

Example of a research question: What factors motivate employees to work in the field of supply chain management?

2.1.1 Set, unit and sample

A statistical population is formed by those arbitrarily selected mass (identical) phenomena that we wish to study. To more precisely define a mass phenomenon, we simultaneously use the following three criteria (Vrečar, 2020):

- factual definition: we answer the question »Who or what are we studying?«;
- spatial definition: we answer the question »Where are we studying?«;
- temporal definition: we answer the question »When are we studying?.

By simultaneously defining all three of the listed criteria, we define our studied population with sufficient precision. For the example presented in Chapter 2.1., the definition of the statistical population would be "employees in the field of supply chain management in Slovenia in 2023". It is highly recommended that the term "all/all" be used before defining the substantive criterion, as this way the statistical population is defined even more precisely and separated from the sample, as will be explained below.

Example of defining a statistical population: (all) employees in the field of supply chain management in Slovenia in 2023.

A statistical set (population) consists of a large number of statistical units. The number of units in a statistical population is denoted by the letter N. Depending on the type of statistical set, the number of units in the population may be known or unknown.

Example of statistical unit definition: (one) employee in the field of supply chain management in Slovenia in 2023.

A sample represents a part of the population, the units of which are selected in order to assess the situation in the population. The reasons for conducting research on a sample and not on the population are mainly hidden by time and financial obstacles, since conducting research on large populations requires a lot of time and financial resources. On the other hand, the characteristics of certain populations, in which it is not possible to come into contact with all the units or the final number of all units is not even known, direct us to conduct research on a sample and not on the population. The number of units in the sample is denoted by the letter n, where it is always the case that the number of units in the sample is less than the number of units in the population, or $n<N$.

When choosing a sample, we must always try to ensure that the latter is ***representative*** (Košmelj, 2007), or that it reflects and conveys all the properties of the statistical population. As an example of a statistical population, we can imagine a cake consisting of three parts: sponge, cream and whipped cream. If we only taste one or two of the three previously presented ingredients, we cannot definitely say that the cake tastes good. Therefore, a piece of cake that is cut in this way does not represent a representative sample. A correctly cut piece of cake from top to bottom, which also includes sponge, cream and whipped cream, is an example of a representative sample, since we can comprehensively judge the taste of the cake based on the tasting of all its parts.

To extract a sample from a statistical population, we always use at least one, previously determined criterion, which we apply sensibly according to our research question and the goals we want to achieve. This can be an additional factual criterion, a location criterion, a time criterion, or a combination of these.

Example of sample definition: employees in the field of supply chain management in Slovenia in 2023 with a company headquarters in the Savinja statistical region.

2.1.2 Variables and parameters

Statistical variables (or just ***variables***) represent the properties of statistical units. We use them to describe our statistical units. They represent the common characteristics of statistical units that we want to study through the research process. The characteristics of the entire statistical population are called ***parameters*** (Vrečar, 2020).

Examples of variables/parameters: gender, level of education, length of service in years, number of employees in the company, etc.

The main characteristic of each variable is that it has its own values, which differ from each other (change). Depending on the type of value it can have, all variables are divided into two large groups, specifically (Bastič, 2006):

- non-numeric or descriptive variables whose values cannot be expressed in numbers;
- numeric variables whose values are expressed in numbers.

Examples of descriptive (non-numeric) variables: gender and level of education, since the values male/female/other and primary school/secondary school/college/university/master's degree/doctorate of science do not have numerical values.

Examples of numerical variables: length of service in years and number of employees in the company, since they can have numerical values from 0 to 40 or from 1 to + infinity.

To make it easier to distinguish between descriptive and numerical variables, we can also use distinctions in the way we think when we observe the entirety of our research problem. When operating with non-numerical variables, we always think in terms of ***proportions*** (e.g., the proportion of female supply chain management employees, the proportion of supply chain management employees with a Master of Science degree, etc.). When operating with numerical variables, we always think in terms of averages (average length of service of supply chain management employees, average number of employees in logistics companies, etc.).

It is important to highlight the fact that numerical variables are higher quality variables from a data processing perspective, as they allow for a greater number of statistical tests to be performed and consequently yield more useful findings in business decision-making. The values of numerical variables can therefore be displayed with a smaller number of descriptive values and consequently change the way of thinking with a focus on shares (for example: the share of employees with an average service period of up to 10 years in the field of supply chain management).

Unlike numerical variables, descriptive variable values cannot be represented by numerical values, as they are lower quality variables. Unfortunately, research often involves errors that occur in the process of coding variable values. For example, the values of the gender variable are assigned numerical codes: female = 1, male = 2, other = 3. It is important to note that the numbers mentioned here are simply a code that replaces the letters »male«, »female« and »other« and are in no way numerical values from which averages could be calculated. »What is the average gender of employees in the field of supply chain management?« is an example of such an unsuitable research question.

2.1.3 Measuring scales

Depending on the method of measurement or the types of values they have, all variables can be divided into four large groups. It should be borne in mind that each of the previously presented types of variables has its own two separate measurement scales (Vrečar, 2020).

Descriptive variables are divided into two groups depending on the measurement scale used, specifically:

- nominal;
- ordinal variables.

Nominal variables represent the lowest quality variables. They allow us to determine whether two values of a variable are the same or different. Ordinal variables allow us to place the values of a variable in order and in this way determine which value is greater and which is smaller. It is necessary to highlight the fact that in no case are these numerical values (greater, smaller or equal), therefore any mathematical operation other than counting is inadmissible in the case of nominal or ordinal variables.

Numerical variables are divided into two groups depending on the measurement scale used, specifically:

- interval or spaced;
- ratio or proportional variables.

Using a spaced measurement scale, we can calculate (numerical) differences between different values of variables. One of the characteristics of these variables is that they often have values that are limited by a minimum and a maximum and contain a specific unit of measurement. Most numerical variables are proportional variables, the main characteristic of which is that we can calculate (numerical) relationships or proportions between them. The value of these variables is not limited in theory, so they can have values from 0 to $\pm \infty$. As already highlighted, numerical variables are variables of higher quality, which allow for the implementation of a larger number of statistical tests and, consequently, higher-quality inference in a business environment.

Example of a nominal variable: location of a warehouse (Celje, Maribor, Ljubljana, Koper).

Example of an ordinal variable: year of construction of a warehouse (1985 < 1989 < 2000).

Example of an interval variable: air temperature in a warehouse (10 C ° in the morning, 8.5 C ° at night). The difference is 1.5 C °.

Example of a proportional variable: stocks of product A001 in warehouse A - 1200 pieces, stocks of product A001 in warehouse B - 2400 pieces. Warehouse B is twice as stocked with product A001 as warehouse A.

If we look at all types of variables as a common unit, we can see that the variables are arranged in order from nominal, which is the lowest quality, to ratio, which is the highest quality. Each subsequent type of variable takes on the properties of the previous ones and has an additional property that improves its quality. A schematic representation of this is presented in Figure 1.

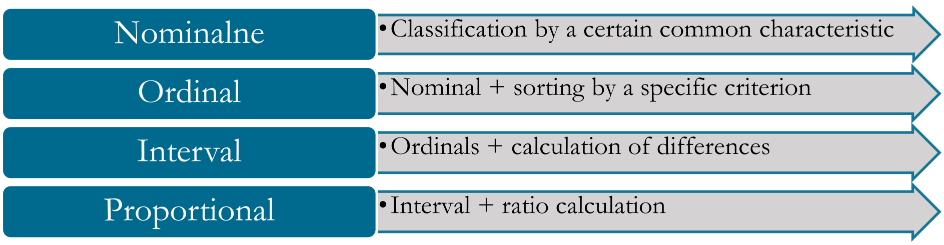

Figure 1: Schematic representation of variable properties according to a measurement scale
Source: own.

2.1.4 Assumptions (hypotheses)

Assumptions or hypotheses are used when we want to present our research question more concretely. In other words, with a hypothesis we concretize our research problem and write it down with a series of clearly formulated statements that can be confirmed or refuted with the help of appropriate data processing and the implementation of an individual statistical test. The main component of all hypotheses are variables.

All assumptions are divided into two large groups (Bastič, 2006):

- null hypotheses, which basically state that there are no (statistically significant) differences, relationships, or influences between variables;
- research hypotheses, which basically state that there are (statistically significant) differences, relationships, and influences.

Null hypotheses are not the subject of research work. When formulating hypotheses in business decision-making and research, we focus on research hypotheses, which are divided into:

- indirect research hypotheses, which are aimed at identifying differences, connections and influences without determining the direction and strength of the influence (it differs, they are related, influences);
- direct research hypotheses, which determine the direction of the differences (greater/smaller) and the strength of the influence connection (positive/negative, strong/medium strong/weak).

Example of a null hypothesis: The quantity of product A001 in stock in warehouses in the Western cohesion region of Slovenia is (statistically significant) ***not different*** from the quantity of products in stock in warehouses in the Eastern cohesion region of Slovenia.

Example of a two-sided research hypothesis: The quantity of product A001 in stock in warehouses in the Western cohesion region of Slovenia is (statistically significant) ***different*** from the quantity of products in stock in warehouses in the Eastern cohesion region of Slovenia.

Example of a one-sided research hypothesis: The quantity of product A001 in stock in warehouses in the Western cohesion region of Slovenia is (statistically significant) ***greater*** than the quantity of products in stock in warehouses in the Eastern cohesion region of Slovenia.

The phrase statistically significant is also given in brackets. Statistical significance refers to the criterion used to assess the significance of the results of a statistical analysis, or in other words, it is the percentage of risk that the conclusion that we will draw based on the risk analysis is not due to the fact that we are stating. Statistical significance is usually based on a certain level of risk, called the significance level, which is chosen before analyzing the data and is expressed in percentages (the most commonly used values are 1% and 5%). In other words, we are talking about a fact or percentage that we could be wrong about. Statistical significance is a mandatory component of all research hypotheses based on statistics, as long as we use it in business decisions as required according to the goals of the individual research (Bastič, 2006).

2.1.5 Primary and secondary data

Depending on the research topic, we distinguish between two groups of data: primary and secondary data (Vrečar, 2020).

Primary data are data that are collected specifically for the purpose of solving a specific (concrete) problem. This is data that requires more time to be spent on its collection and editing phases, but its main advantage is that it can be adapted to the specific topic of a specific research problem. The following methods are most often used to collect primary data:

– **Survey**: The key part of this method is the composition of the survey questionnaire or the preparation of a group of questions and answers that the respondent chooses themself. There are several ways to conduct a survey (in person, by phone, by regular mail or by e-mail). The survey questionnaire can be prepared in a classic manner or with the help of various programs that enable sharing of survey questionnaires online (example: Forms in GoogleDrive, Microsoft Forms, 1ka, etc.). Surveys mostly contain closed questions (pre-selected answers that the respondent chooses, yes/no, numerical measurement scales), but they can also contain open questions, often to a smaller extent. The main purpose of the survey is to check the strength of an individual option, or we expect respondents to confirm our assumption as to which option the respondents could select most frequently.

- **Interview**: The key part of this method is a conversation in which the interviewee answers questions that can be open or closed. Open-ended questions are mostly used in interviews, as we expect the interviewees to answer questions such as: what, how, what do you think, etc? The interview is often the starting point for the brainstorming method, the search for new solutions, or the basis for preparing a survey questionnaire in cases where we want to combine qualitative and quantitative research methods.
- **Observational study**: The key part of this method is observation, which most often refers to the observation of human behavior. However, we can also observe the product range, individual elements of the company, buildings, etc. The starting point for conducting an observational study is an observation sheet, which contains a combination of open and closed questions.

Unlike primary data, secondary data is already collected and stored in a specific location, which means that when using secondary research data, we save time and resources required for their acquisition. Their weakness is that they are quite inflexible, sometimes inadequately broken down, geographically or temporally inappropriate for the needs of our research, etc. Secondary data is therefore already collected from institutional providers of secondary data, such as: AJPES, Business Register of Slovenia, Statistical Office of the Republic of Slovenia, EUROSTAT, OECD, United Nations and similar.

Regardless of the type of data (primary or secondary data) or the method of data collection (survey, interview, observational study), the raw data collected must be stored in a systematic manner. For this purpose, we use a **data collection framework**. This is a structure or plan that we use when collecting information for a research project or study. It includes certain guidelines, methods and procedures that help to systematically obtain data in a way that ensures the reliability, accuracy and comparability of results. The data collection framework also includes decisions about how the data will be collected, what types of data will be captured, who will be involved in the research and how the data will be analyzed. This framework is essential for ensuring the quality and relevance of the data obtained in the research process.

In practice, a data collection framework means a survey questionnaire, interview, observation sheet or spreadsheet into which the raw data will be entered. After the data collection process is completed, the final result of this phase is a spreadsheet with raw data, which needs to be refined and prepared for the implementation of appropriate statistical analyses.

2.2.6 Tabular and graphical display of data

Data is presented in two basic ways: tabular and graphical. Both methods of presenting data must be appropriately equipped or contain all the necessary elements in order to be sufficiently transparent for the reader.

Each table contains the following mandatory elements (Vrečar, 2020):

- ***table title***, which must answer four questions that in Slovene begin with the letter K (What, When, Where and How);
- ***table header***, which includes column names;
- ***table header***, which includes row names in two-dimensional tables (e.g. inventory status by product and by month);
- ***rows and columns with data*** that form ***fields (cells)*** at the intersection;
- ***summary rows and summary columns***, which display summarized data, such as total stocks by product, total stocks by month, and the like;
- ***data source,*** which can be external or internal (own);
- ***additional methodological explanations,*** what the element represents, which is used as needed and with which we provide, for example, the calculation methodology, any problems with the data source, any additional explanations regarding the data, etc.

Table 1 provides a good example of a table that covers all the required elements.

Table 1: Number of products in stock by product serial number and month for the company Zaloga d. o. o for the third quarter of 2023

Product serial number	August	September	October	Total
A001	30	70	40	140
A002	50	30	50	130
B001	100	44	45	189
B002	80	25	60	165
C001	20	73	80	173
Total	280	242	275	797

Source: own research

We use ***graphs*** to highlight the data that is most important from the perspective of our research. As with tables, the rule is that graphs must be equipped with mandatory elements, such as: title, source and, if necessary, additional methodological explanations. Vrečar (2020) lists some basic rules for the correct presentation of data using graphs:

- it is important to ensure correctness of data display (e.g. if one phenomenon is twice as large as another, this must be displayed correctly);
- it is not sensible to create a chart if there is very little or very much data or if the phenomenon is very stable (does not vary much);
- when using bar charts, make sure that the starting point of the y-axis is at the value 0;
- it is very important to choose an appropriate chart, which, depending on the needs of the research, is divided into ***simple*** (bar charts, line charts, pictograms, cartograms) and ***analytical***, which are considered tools for statistical analysis (structural circle, histogram, polygon).

Figure 2 presents a good example of a simple chart that includes all the necessary elements.

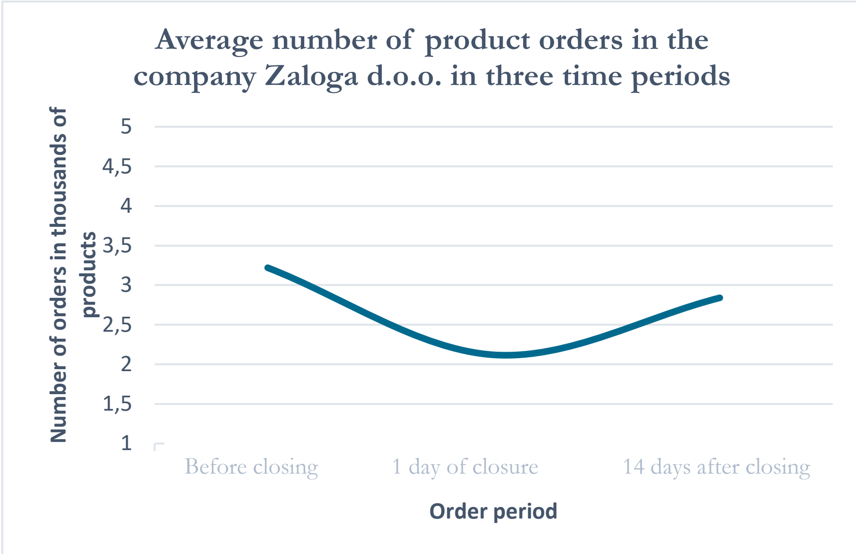

Figure 2: Graph example
Source: own.

2.2.7 Choosing the appropriate statistical analysis

The choice of the appropriate static analysis depends on the way in which we have formulated our research hypothesis. Basically, we distinguish two large groups of static analyses (Bastič, 2006):

– ***descriptive statistics***, which focuses on describing and summarizing the data collected. It uses various methods such as: means, ranges and graphical displays to reveal the basic properties of the data set. The goal is to present information in an understandable way, which helps to form a clear insight into the characteristics and distribution of the data, without attempting to establish causal or statistically significant relationships between variables;
– ***inferential statistics***, which deals with drawing conclusions and generalizing from a sample of data to the entire population. This attempts to answer questions about the population based on statistical analyses of the sample, using approaches such as confidence intervals and various statistical tests.

3 Conclusion

Based on the content presented in this chapter, we can conclude that the basic concepts presented are of vital importance and understanding them plays a key role in the process of correct business decision-making. A basic understanding of data analysis concepts proves to be a key tool that allows companies to effectively utilize the vast amounts of data they collect. With in-depth knowledge of these basic concepts, we enable organizations to more successfully manage information, which leads to better and more thoughtful business decisions.

In addition, this chapter has highlighted how a basic understanding of analytical methods and approaches can contribute to improving decision-making processes in an organization. With the correct interpretation of data and the use of analytical tools, decisions become more focused and goal-oriented. This is reflected in more effective business decision-making strategies, which are crucial for the long-term success of a company.

Ultimately, based on the content presented, we can conclude that a fundamental understanding of the basic concepts of data analysis is key to achieving a competitive advantage in the market. Organizations that strive for excellence in data-driven business decision-making will be better prepared for the challenges and opportunities brought by the modern business environment, such as the green transition, digitalization, and ensuring long-term resilience, through a deeper understanding and appropriate application of these concepts.

References

Bastič, M. (2006). *Metode raziskovanja*. Univerza v Mariboru: Ekonomsko-poslovna fakulteta. https://www.yumpu.com/xx/document/view/19079362/metode-raziskovanja-majda-bastic-shranisi

Košmelj, K. (2007). *Uporabna statistika* (2nd ed.). Biotehniška fakulteta. https://repozitorij.uni-lj.si/Dokument.php?id=132497

Vrečar, P. (2020). *Poslovna informatika s statistiko: Študijsko gradivo z vajami* (8th ed.). Višja strokovna šola za gostinstvo in turizem.