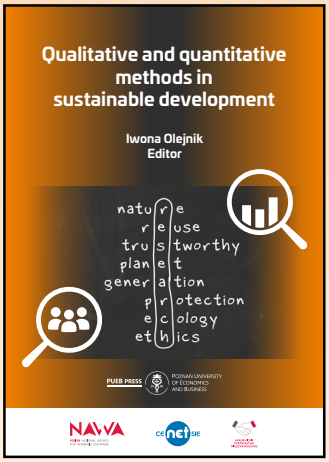

Qualitative and quantitative methods in sustainable development

Iwona Olejnik
Editor

NAWA POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

cenet SIE

AKADEMICKIE PARTNERSTWA MIĘDZYNARODOWE

eISBN 978-83-8211-072-2

https://doi.org/10.18559/978-83-8211-072-2

PUEB PRESS

POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

Qualitative and quantitative methods in sustainable development, pp. 41-79
https://doi.org/10.18559/978-83-8211-072-2/02

2.

QUANTITATIVE METHODS

Sylwester Białowąs Blaženka Knežević, Iwona Olejnik, Magdalena Stefańska
Poznań University of Economics and Business

Blaženka Knežević
University of Zagreb

Iwona Olejnik
Poznań University of Economics and Business

Magdalena Stefańska
Poznań University of Economics and Business

Abstract: The main goal of the chapter is to present the basics of survey research that can be used in analyzes of sustainable development.

The first part presents the measurement levels. The basic characteristic of every variable is its level of measurement. It implies the following analysis and available techniques. This part introduces four levels of measurements: nominal, ordinal, interval and ratio, showing their characteristics and examples. Then the focus is on the implications of a given level of measurement on the possibilities of the statistical analysis.

The aim of the second chapter is to explain the foundations of preparing a questionnaire for the research on the issues related to sustainable development. An example of an organic food questionnaire is also provided.

The third part presents considerations necessary for the sampling process. The main goal is to present the basic methods of calculating the minimum sample size, as well as the methods of its selection. This section presents the arguments for conducting the study on a sample rather than on the entire population, and also several formulas enabling the calculation of the minimum sample size. A discussion of the most important methods of selecting respondents to the sample—both random and non-random, can also be found here.

The last two parts of this chapter, describe the ways of presenting the results of quantitative research. They describe first view of the variables including frequency distribution with charts, central tendency measures and cross-tabulation. Finally, the methods of presenting research results obtained on the basis of the Likert scale and other examples of data visualization schemes are presented.

Keywords: data presenting, levels of measurement, questionnaire, sampling.

2.1. Levels of measurement

2.1.1. Introduction

Thinking about measurement we usually connect the idea with physical items. We can easily imagine measuring width, length or weight. We have the proper tools to say that the table is 1,4 meters long and the mobile phone weights 120 grams. It may look different in economics, still we measure physical objects, but lots of the measurements are related to non-physical concepts, like happiness, financial literacy, attitudes toward clean energy or opinions about waste management.

The results of the measurements can be compared more or less precisely. Talking about the country of origin we can just tell if it is same or different. But when we measure salaries we can easily say that three thousand euro is not the same as two thousand euro; the difference is one thousand euro, and the three thousand is 1,5 times more than two thousand euro. So, the comparison is full of new details. How can we compare the measurement results and what statistic techniques can we use considering the level of measurement?

There are four levels of measurement:

1) nominal,
2) ordinal,
3) interval,
4) ratio.

2.1.2. Nominal, ordinal, interval and ratio levels of measurement

The nominal level of measurement is the lowest and simplest one. On this level we can assess the observation in terms of the category. Every observation may be assessed as belonging to one of the categories. It may not be attributed to more than one category.

According to such assessment we can confirm, if the elements belong to the same category, or if they belong to different categories.

Example of the nominal level of measurement:
- *Gender: male, female*
- *Believing in human driven climate change: yes, no*
- *Believing that one of the countries named below will be the first one to reach carbon neutrality (net carbon dioxide emission): Bulgaria, Croatia, Czech Republic, Hungary, Poland, Slovakia, Ukraine*
- *Sorting waste in your household: yes, no*

Usually, we use numbers for coding the categories, and in the case of nominal level for measurement, the numbers have no meaning. Using codes like:

1. *Bulgaria,*
2. *Croatia,*
3. *Czech Republic,*
4. *Hungary,*
5. *Poland,*
6. *Slovakia,*
7. *Ukraine*,
8. *Other*,

or

1. *Ukraine,*
2. *Slovakia,*
3. *Poland,*
4. *Hungary,*
5. *Czech Republic,*
6. *Croatia,*
7. *Bulgaria,*
8. *Other*

gives the same possibilities in interpretation.

What should remain constant is the last position of "other". "Other" is used very often if we cannot name all possible answers.

Data from a nominal scale should not be used for arithmetical calculations, it would be meaningless. It does not make sense to replace Czech Republic (represented by value 3) by combination of added Bulgaria (1) and Croatia (2). And there is no point in presenting the arithmetic mean from the countries from the above example.

If the variable has only two possible categories we call it a dichotomous variable. The distinction is important, as some statistical methods can be used for dichotomous variables and cannot be used for other nominal variables. For example, the question about believing in human driven climate change, with two possible answers (yes, no) is a dichotomous variable.

The ordinal level of measurement offers us more than nominal data. Using the ordinal scale, we not only compare if the elements are the same or different, but we can also order them according to the intensity or time. However, the data tell us nothing about the differences between the values.

Imagine the three households A, B and C. The households compare produced waste. Household B produced the most, whereas household C produced the least. However, we know nothing about the differences. Was household A more similar in waste production to B or C? Maybe it produced almost the same number of waste bins as B with a huge difference to C? This could also be completely different.

Example of ordinal level of measurement:
- *Performance in waste production (first, second, third)*
- *Liking this e-book (dislike very much, rather dislike, rather like, like very much)*
- *Supporting state investments in coal mining (fully against, rather against, rather accepting, fully accepting)*
- *Attitude toward sorting waste: very negative, rather negative, rather positive, very positive*

The numbers are important and have to be assigned to the verbal levels in ascending or descending order.

Interval level of measurement is even more useful. We not only compare if the elements are the same or different, or not only can we order them, but we are able to measure the distances between them. Data are interval, if the differences can be measured in units, and the units have equal intervals between them. The interval variables can have negative values.

Example of interval level of measurement:
- *Bank account balance (e.g., –270 EUR, 50 EUR, 70 000 EUR)*
- *Temperature of the sewage in main sewer collector in Celsius (e.g., 17, 21)*
- *The year you have started sorting waste in your household: (e.g., 2001, 2002, 2003…)*

We can calculate that Samantha's household (1995) started sorting waste six years earlier than Bob's household (2001).

Ratio level of measurement goes a step further. Data collected on this level allow for comparing elements, ordering them and measuring the distances between them. And additionally, the ratio level allows for calculating meaningful ratios of values along the scale. This is possible only if the scale has true and meaningful zero point. The zero value indicates in this case, that there is none of that variable. Of course, the ratio level variables cannot have negative values.

Example of ratio level of measurement:
- *Number of bumblebees visiting a given flower in one hour (e.g., 0, 5, 35)*
- *Weight of the waste produced in one month by the household (15 kg, 45 kg, 70 kg)*
- *Tons of unsorted waste reported by municipality of your city in 2020: (e.g., 3, 250, 781)*

2.1.3. Measurement levels and data analysis methods

The level of measurement has serious impact on how we can analyze them. You will know more about it in the next part of the book. Table 2.1 shows the central tendency measures and basic calculations available for given levels.

Table 2.1. Levels of measurement and basic calculations

	Nominal	Ordinal	Interval	Ratio
Mode, frequency distribution	+	+	+	+
Median, presenting order		+	+	+
Calculating mean, adding, subtracting			+	+
Calculating ratio				+

Source: Own elaboration.

Levels of measurement imply the permissibility of statistical techniques. When forming hypotheses and designing the questionnaire for the primary research we always have to plan the method of future analysis and check its feasibility.

Table 2.2. Permissible descriptive and inferential statistics for certain levels of measurements

	Descriptive	Inferential
Nominal	percentage, mode	chi-square, binominal test
Ordinal	percentile (e.g., median)	rank-order correlation, Friedman ANOVA
Interval	range, mean, standard deviation	product-moment correlations, *t*-test, ANOVA, regression, factor analysis
Ratio	geometric mean, harmonic mean	coefficient of variation

Source: Own elaboration based on: (Malhotra & Birks, 2003).

2.2. Questionnaire design

2.2.1. Stages in creating a questionnaire

The questionnaire is the basic research tool used in the primary quantitative research, both in the interview methods (in which the questions are asked, and answers are written by the interviewer) and in the questionnaire (in which the respondent reads the questions and writes down the answers himself). It is an indispensable research tool in many types of research: door-to-door interviews, executive interviews, mall-intercept interviews, telephone interviews, self-administered questionnaires, and mail surveys

(McDaniels & Gates, 2018). The questionnaire is a set of written questions formulated to elicit the desired answers related to the examined problem (Neuman, 2014).

A properly constructed questionnaire should meet many rules concerning both the content of the questions, the way they are formulated, the scaling of the answers, the order of the questions, and the graphic design (Churchill & Iacobucci, 2018). The questionnaire should encourage the respondent to participate in the survey and make the survey interesting for them. At the same time, it is designed to facilitate the work of the researcher (especially the interviewer) through instructions on how to ask questions and write down the answers.

The following stages of building the questionnaire can be distinguished:

1) specification of the purpose and subject of the questionnaire,
2) formulation of an initial list of questions,
3) initial scaling of the response,
4) initial check of the questionnaire,
5) construction of the sample questionnaire,
6) pilotage and possible modifications of the questionnaire,
7) preparation of the final version of the questionnaire.

When starting to build the questionnaire, one should remember about the assumptions of the research project, in particular, about defining who the respondents will be (their level of education, age, etc.) and what their knowledge is in relation to the subject covered by the research (general or specialist?)—it determines form of asked questions and the use of specific phrases.

The construction of the questionnaire, including the number of questions and their form, depends on the research method. The form will look slightly different in the interview method than in the survey method. These methods differ from each other, inter alia, in the form of contact with the respondent (direct, indirect) and the number of questions asked and answers received. An important determinant of the selection of the type and construction of the questionnaire is also the data that is to be collected—their scope and details, as well as the costs to be incurred, time and planned methods of analysing the obtained data.

Regardless of the intended research method and the type of questionnaire, certain elements are present in each of them. The questionnaire therefore consists of three basic parts:

1) introductory part—the header,
2) the main part—containing questions aimed at solving a research problem,
3) respondent's particulars—personal information (data about the respondent).

The introductory part—the header of the questionnaire—should include such elements as: the name and contact details of the organizer of the study, the title of the questionnaire, a short description of the purpose, nature and scope of the study, and possible indication of benefits for the respondent resulting from the

participation in the study, ensuring that the survey is anonymous (optional), the date of return (in the case of forms returned by post), the name and surname of the interviewer and, possibly, information on how to complete the questionnaire.

2.2.2. Types of questions in the questionnaire

The main part of the questionnaire consists of the subject-related questions concerning the research problem. There should be enough questions to reflect the entire substantive scope of the planned research, and thus to enable the verification of the research hypotheses set out in the project, including the following categories of questions (Neuman, 2014, p. 317):

- behaviour (*How frequently do you ....., When did you last .....*),
- attitudes, beliefs, opinion (*What is the biggest problem facing ....., What do you think about .....*),
- expectations (*Do you plan to buy ..... in the next 3 months*),
- self-classification (*Do you consider yourself to be .....*),
- knowledge (*About what percentage of ..... It is legal to own .....*).

The questions in the questionnaire can be open or closed. Open-ended questions do not have any given options, i.e., the respondent formulates the answers in a completely arbitrary way. They are used when it is difficult to predict the answers of the respondents, when the number of these answers may be too high, or when they concern difficult issues, e.g.:

> How should food waste be prevented? Please develop your answer / write down your suggestions .....
> .....
> Please, finish the statement: „In second hands the most irritating thing is ....., because .....”

Closed-ended questions (with a scale) provide for the selection of answers from a prepared set of options predicted and determined in advance by researchers, i.e., from the so-called cafeteria. These questions are also called scaled questions. It should be remembered that the consequence of the selection of a specific scale are—after the research is conducted—the methods of processing and analysing the data obtained using them (Aczel, 2009). As a result, the so-called one-dimensional and multi-dimensional scales are applied (Neuman, 2014).

An example of a one-dimensional closed question is, e.g.:

> Do you think that biodegradable waste should be sorted separately? *Please underline 1 answer*: a) definitely yes, b) rather yes, c) rather not, d) definitely not.

On the other hand, when constructing complex questions, the following scales can be used:

- ranks: they give the opportunity to rank the most important categories for the respondent (things, attitudes, etc.) according to a given criterion,
- rank scale of summed up grades: it gives the possibility to rank the most important things, attitudes, etc. for the respondent by dividing a certain pool of points (usually 10 or 100),
- comparative scale of summarized features: the respondent assigns the listed features or objects, according to their preferences, numerical ratings so that they give the required sum (usually 10 or 100),
- semantic differential (semantic): it allows to obtain the respondents' opinions about a given object in the range between two opposite attitudes,
- positional: it consists of many individual words or phrases in any order, which are assessed by the respondents according to their preferences using the same ordinal scales,
- Stapel's: it was designed for situations in which some features of the tested objects do not have obvious opposites,
- Likert: it allows you to specify the degree of acceptance of a given statement or statements (from 1—strongly disagree, to 5 (or 7)—strongly agree).

As mentioned before, a properly constructed questionnaire should meet many rules regarding the structure of questions. First of all, "the language of the questionnaire should approximate the level of understanding of the respondents (Bougie & Sekaran, 2020, p. 147). So, they should:

- be short and simple (with some exceptions),
- avoid jargon, slang and abbreviations,
- be unambiguous in terms of the vocabulary used—understandable in the same way by each potential respondent,
- not suggest an answer; avoid emotional language and prestige bias,
- explain a priori specialist words and phrases,
- have comprehensive scales—containing all potential answer options,
- be formulated using homogeneous grammatical forms.

Questions should also be placed on the form in the correct order, logical from the respondent's point of view. In the case of research on consumer behaviour and attitudes in relation to various products and services:

- first, questions about the awareness of their existence are placed,
- then questions about behaviours, e.g., in terms of the volume of purchases of a given product and competing products, frequency of purchase, ownership, decision-making process,
- finally, questions about opinions and motives, as well as anticipating future purchases.

In addition, in the case of a questionnaire dealing with various issues, it is worth dividing the questions into sections—according to the thematic threads discussed, using additional spaces and subheadings. Different fonts for the content of the questions and for instructions for the interviewer or respondent are also worth using.

In the last part of the questionnaire—respondent's particulars (respondents' personal questions) are included. There are subjective questions characterizing the research unit. The lack of these part of questionnaire is a fatal error of the study, as it makes it impossible to assess its representativeness and to conduct segmentation. In the case of surveys conducted among individuals and households, this part includes questions about gender, age, occupation, number of people in the household, number of children, place of residence, income and others, while in the case of surveys conducted among enterprises these are questions about the industry, form of ownership, number of employees, location, period of the company's existence, scope of activity, position of the person giving the answer and others.

At the end of the questionnaire, you can also write acknowledgements to the respondent for participating in the survey, as well as any organizational information (by when and where to send the form). Additionally, if the research is performed using the interview method, the last element is the part for the interviewer, in which they write down: the date, time of the beginning and end of the interview, information about the interview conditions (respondent's reactions, presence of third parties, etc.), information about the questionnaire (e.g., which questions were difficult for the respondent, which were sensitive, etc.) and his / her signature.

When the questionnaire seems ready to start your research, it is worth re-evaluating it and answering the questions in Table 2.3.

Table 2.3. Questionnaire assessment

– What do we want to get with the question?
– Will the respondent know the answer to a given question? What answers will we get? What is one solution to the research problem—what?
– Are options of answers adequate to the question posed?
– Is it possible to make the choice of the question easier (without losing information)?
– Why is this question situated here?
– What is the expected method of analysis?
– What is the reliability of the data obtained?
– Have we not asked about it before?
– Are there any errors in the language notation?
– Should not there be a filtering question in advance?

Own elaboration.

Before the actual examination, a so-called pilot study is conducted, i.e., a preliminary test that verifies the correctness of the prepared form. Usually, it is performed on a small pilot sample, e.g., 5–15 people, but depending on the complexity

of the study, it may cover even a larger number of respondents. Its objectives are to determine the respondents' reactions to the objectives of the survey, to find out whether the questions included in the questionnaire pose difficulties for the respondents, and then to make the necessary changes in the questionnaire as well as write instructions for respondents and / or interviewers.

2.2.3. Organic food as a form of sustainable consumption: case study

The aim of the research conducted with the use of the direct interview method was to identify the attitude of young parents to organic food as a form of sustainable consumption. Before the questionnaire was created, the following hypotheses[1] had been formulated (they were then reflected in the prepared form):

H1: Young parents are aware of the meaning of the organic food consumption concept

H2: Young parents perceived limited access to organic products as the most common reason among barriers of the organic food consumption

H3: Young parents in bigger families implement more activities connected with the buying organic food.

The researcher started to design the questionnaire and prepared its draft (check below).

The data of institution that conducted the research

How and why is organic food important?

We would like to invite you—young parents—to share your opinions about the attitude to organic food consumption. Your answers are completely anonymous and will only be analysed in the form of aggregate results and not be used for marketing or any other purposes non-related to this survey. The survey takes around 15 minutes to complete.

1. Who buys organic food products in your household?
   a. I buy
   b. My partner
   c. Both
   d. Other person, *who* …….

[1] Hypotheses are „predictions the researcher makes about the expected outcomes of relationships among variables. They are numeric estimates of population values based on data collected from samples" (Creswell & Creswell, 2018, p. 136).

2. What organic food products do you usually buy: .....
.....
.....

3. From what sources do you obtain information on organic food products available in the market? *Please sort by importance and enter the numbers in the boxes, from left to right*

a. My partner	b. Friends	c. Daily press
d. Internet	e. Family	f. Shop leaflets
g. Others		

1… 2… 3… 4… 5… 6… 7…

4. What is the most important for you when you buy food *(please select the 3 most important factors from the following, and rank them by entering the response codes into the boxes, treating the first box from the left as the most important)*:

a. Price	b. Type of packaging (recycled or not)	c. Production place
d. Ingredients	e. Promotion	f. Size of packaging
g. Brand	h. Price reductions for a given product	i. Type of packaging (can be recycled or not)
j. Expiration date	k. Recommendation of the staff	l. Others……

1..... 2..... 3.....

5. How important are the following criteria for you when choosing food products? *Please divide the 100 points between these factors in such a way as to give the most points to the one you consider most important, and correspondingly less points to the less important ones. If you deem any factor to be invalid, please do not score it.*

Criteria	0–100
store location	
package size	
price	
manufacturer	
method of production	
country of origin	
ingredients used in production	
caloric content	
nutrient content	
others (please write your answer)	
Total:	100

6. Please divide 100 points between each of the following four pairs of different types of stores according to your preferences
   supermarket (.....) + convenience store (.....) = 100
   convenience store (.....) + local market (.....) = 100
   local market (.....) + discount stores (.....) = 100
   discount stores (.....) + supermarket (.....) = 100
7. Please mark your grades that best reflect your opinion on the mentioned characteristics of eco-chocolate. *Please insert an X*

tasty						distasteful
cheap						expensive
smells good						smells bad
nicely packaged						poorly packaged
very healthy						less healthy

8. How do you rate the features of our supermarket as listed below? *Please insert an X*

	I like it very much	**I rather like it**	**I don't really like it**	**I don't like it at all**
Parking places				
Location				
Appearance of the building				
Interior decor				
Wide of assortment				
Facilities for families with children				
Services provided by employees				
Atmosphere				
Level of prices				
Attractiveness of promotions				

9. A number of different views on different purchasing issues are presented below. For each of these opinions, please tell me to what extent you personally agree with them. *Using the scales from 1 to 5, where 1 means that you completely disagree with the opinion, 2: rather disagree, 3: neither agree nor disagree, 4: rather agree, and 5: that you fully agree, please respond to the following statements:*

	1–5
I always try hard to reduce misuse of goods and services (e.g., I switch off light and fan when I am not in the room)	
I sort food packaging to recycle it	
I avoid being extravagant in my purchase	
I avoid overusing / consumption of goods and services (e.g., take print only when needed)	
I choose to buy product(s) with biodegradable container or packaging	
I don't like to waste food or beverages	
I recycle my old stuffs in every possible way (e.g., distribute old clothes among needy people)	
I always search for products with symbols of organic food	
Being a parent made me more sensitive about food we consume	
I don't buy food from unknown sources	
I search for new organic products	
We buy organic food branded only by producers, not stores	
I believe in organic food	
I don't buy more than we need in our household	
I don't buy ready to eat only organic food	
I usually buy not processed organic food	
I put a lot of energy to find organic food I need	
I put a lot of energy to learn more about product ingredients	

10. The monthly percentage of expenditure on organic food in your household is around: ……%

And finally, a few words about your household—characteristics

1. Gender:
   a. female
   b. male.
2. Age ……
3. Job/occupation ……
4. Which of the following sentences describe the best the financial situation of your household. Our household income:
   a. allows us to purchase luxury goods,
   b. allows us to set a small financial surplus,
   c. allows us to satisfy everyday needs,
   d. is not enough, we spend savings,
   e. is not enough, we are indebted.
5. Are there children in your household? How many? ……
6. Place of living: ……..……

Thank you for your answers!

2.3. Population and sample

2.3.1. Difference between population and sample

When a researcher wants to explain some of the characteristics, trends or interdependencies in a certain group of people, animals, objects, events, countries or phenomena, the group is called a population. Thus, a **population** is the entire group about which conclusions from the research are drawn (Barrow, 2017).

When we collect data, we can decide to collect it from an entire population or from a sample. Usually, population size and geographical spread are problems that can significantly lower cost-efficiency of the research and that can increase complexity of the research management. If a population is large or difficult to reach geographical area, a researcher can use a sample to estimate or test hypotheses about population data. A **sample** is a smaller, specific group extracted from population that will be used to represent population in a research (McDaniel & Gates, 2018). Using samples allows researchers to conduct their studies more easily, and in a faster and more cost-efficient way.

Regarding collected data, we distinguish a **parameter** as a measure that describes the entire population and a **statistic** that describes the sample. Estimation or hypothesis testing is used to describe how a statistic differs from the population parameter. The difference between population parameter and a sample statistic is called a **sampling error**. The larger the sample, the lower the sampling error (Dean & Illowsky, 2013).

2.3.2. Determining sample size

Sampling is the process of obtaining information from a subset of a larger group. If we select an appropriate number of people for the research conducted on a sample, we can take results and project them to the larger group (Levy & Lemeshow, 2008). The sample should be a true miniature of the population, which interests us considering the research problem (Acharya, Prakash, Saxena, & Nigam, 2013).

The sample size is always smaller than the size of a population. There are some recommendations on how to determine how large a sample should be in order to ensure accuracy of research conclusions on targeted population (Kasiulevičius, Šapoka, & Filipavičiūtė, 2006).

Sample size depends on the following:

1) population size (N),
2) prevalence (p),
3) margin of error (e),
4) sampling confidence level or its z-score (z).

Population size (N) is the total number of items in targeted (researched) population.

Prevalence (p) is a proportion of a population that has a specific characteristic in a given time period. Prevalence can be used from previous survey results or observed by running a small pilot survey. If it is unknown, 0.5 can be used in calculation and it will give the largest possible sample size.

Margin of error (e) is a percentage that describes how close we can expect a survey result to be relative to the real population value.

Sampling confidence level shows reliability of the research; it is expressed as a percentage, which shows the level of certainty regarding how accurately a sample reflects the population within a chosen confidence interval.

In practice, the sample size may also depend on factors such as, e.g., the spatial scope and subject matter of the research, research method and technique, as well as the budget available for the research.

For the purpose of sample size calculation, sampling confidence level is converted into own z-score that is used in the formula by applying the following table of values:

Table 2.4. Confidence level and z-score

Confidence level (%)	z-score (±)
70.0	1.04
75.0	1.15
80.0	1.28
85.0	1.44
90.0	1.65
95.0	1.96
99.0	2.58
99.9	3.29

Source: Own elaboration.

One of recommended formulas (1) to calculate the size of the sample is:

Formula (1)

$$\text{Sample size, } n = N * \frac{\frac{Z^2 * p * (1-p)}{e^2}}{[N - 1 + \frac{Z^2 * p * (1-p)}{e^2}]}$$

The larger the population, the lower the targeted margin of error, and the higher the desired sampling confidence level, the larger the size of the sample. In addition, the accuracy of the research results will increase if we use a larger sample.

When the feature differentiating the respondents' answers (and then taken into account when assessing the representativeness of the sample) is a quantitative one (e.g., household income, expenses for a product), and the population on which we conduct the survey is infinite, the sampling will be independent (with return) and when the confidence interval is assumed based on the sample mean and the sample standard deviation (δ), the formula (2) can be used.

Formula (2)

$$n = \frac{Z_\alpha^2 \delta^2}{e^2}$$

But if the general population is formed by a specific number of N elements, and the distribution of the studied trait x in the population is close to normal, we can apply the following formula:

Formula (3)

$$n = \frac{\delta^2}{\frac{e^2}{Z_\alpha^2} + \frac{\delta^2}{N}}$$

Examples:

1. We want to research how many Internet users in Croatia search for cosmetic products online and we want to analyze what kind of information is the most important.
2. We want to research how retail companies in Croatia change business practices in time of pandemics.
3. We want to estimate the level of food waste and analyze sustainable consumer behavior patterns in rural parts of Poland.
4. We want to analyze information structure of advertisements published at our portal.

Table 2.5. Types and examples of probability sampling methods

Number of research (given examples)	Population description	Population size	Sample determinants (desired values)	Calculated sample size (according to Formula 1)
1	Internet users in Croatia	3 787 838 (Source: Eurostat data for 2019)	margin of error 5% confidence level 95%	385
2	number of active companies in retail industry in Croatia	49 120 (Source: Croatian Statistical Bureau)	margin of error 3% confidence level 95%	1 045
3	number of people living in rural areas in Poland	15 064 972 (Source: https://www.worldometers.info/demographics/poland-demographics/)	margin of error 3% confidence level 90%	757
4	number of advertisements in the last month	1000 (Source: own database of a company)	margin of error 5% confidence level 90%	177

Practical note: There are numerous sample size calculators available online for free (use)! Google and try them!

Source: Own elaboration.

2.3.3. Sampling method

There are two basic methods to select a sample from a population:

a) probability sampling,
b) non-probability sampling.

In **probability sampling** each member of population has a chance to be randomly selected as a member of the sample. This sampling method reduces the risk of sampling bias and gives a better ground for generalization of findings to the whole population. It is widely used in quantitative research. Probability sampling methods are simple random sampling, systematic sampling stratified sampling and cluster sampling. The types and examples of probability sampling are described in Table 2.6.

Table 2.6. Types and examples of probability sampling methods

Method	Characteristic	Examples
Simple random sampling	Each member of the population has an equal chance to become a member of the sample. Sampling takes into account the entire population.	*We want to research opinions of employees on working conditions in company A. The company has 500 employees, calculated sample size is 100. We use the database of employees and by random number generator we generate 100 random numbers to select employees with such ID from the database.*

Method	Characteristic	Examples
		We want to ask questions concerning opinions about local election preferences in some city, calculated sample size is 1,000. We use the register of inhabitants and randomly choose 1,000 people as targeted respondents of the survey.
Systematic sampling	Every member of the population is listed with a number, and we use some interval to select members of the sample.	*We want to research the opinions of employees on working conditions in company A. The company has 500 employees, calculated sample size is 100. We assign numbers from 1 to 500 to each employee and then we select every fifth employee as a sample member. The persons with numbers—5, 10, 15, 20 etc. are sample members.*
Stratified sampling	The population is divided into subpopulation (strata) according to some characteristics and then based on proportion of the characteristics in population, we select a similar number of persons with particular characteristics for our sample.	*We want to observe consumer behavior in rural part of Poland. We check the proportion of females, the proportion of young / old, the proportions according to average income in the sample. Then we divide the population in sub-groups according to gender, age and income. Then we apply random sampling method or systematic sampling method to select a certain number of people from each subgroup we have created. The aim is to ensure same sub-group proportions in the sample. If there is 10% of young females with high incomes in population, then we want to have 10% of them in our sample.*
Cluster sampling	The population is divided into subgroups and each subgroup has similar characteristic as the entire population. Then we randomly choose a subgroup or subgroups to represent the population.	*A retail company has 90 stores operating in various cities across country. In each store, there is a similar organizational structure and there is the same number of middle managers with similar structure regarding education and experience. So, if we want to perform a research on management attitudes towards business ethics in this company, we can randomly choose 5 stores and test managers for opinions.*

Source: Own elaboration.

Non-probability sampling means that members of a sample are selected by using some specific criteria. In this way, the sample is more convenient and accessibility to sample members is facilitated. Thus, non-probability sampling provides a cheaper way to conduct research. Anyhow, due to non-random selection of sample members, validity of generalization of results to an entire population is questionable. Non-probability sampling methods include convenience sampling, voluntary response sampling, purposive sampling and snowball sampling (Etikan, Musa, & Alkassim, 2016). Such sampling methods are often used for qualitative research when the purpose of research is to understand a small, specific or under-researched population. The types and examples of non-probability sampling are described in Table 2.7.

Table 2.7. Types and examples of non-sampling methods

Method	Characteristic	Examples
Convenience sampling	The sample includes persons who are the most accessible to researcher.	*Hotel management wants to examine loyalty of guests, so they ask their guests to fill in the questionnaire on check in or on check out.* *We want to examine consumer ethics of young adults. Therefore, the researcher asks students at the University to participate in the sample.*
Voluntary response sampling	It is based on easy access to respondents, but the researcher is not approaching respondents directly. The researcher places a call for the research and people volunteer to participate in the sample.	*We want to examine the attitudes of young people towards the quality of nightlife in the city. We place a call for research participation at local newspapers, on official web page of the city, and on social networks. Then young people voluntarily answer to the call. With such calls, we have to be aware that voluntary participation can mean that a certain person has a stronger opinion (positive or negative) towards researched topic, thus research can be highly biased.*
Purposive sampling	The members of the sample are chosen based on their characteristic or expertize as the most useful for achieving the research purpose and goal.	*We want to examine the expectation of digitalization in 10 largest retail companies in our country. We make contacts and organize interviews with 20 top managers in those companies.* *We want to research how mothers with small children are satisfied with HR policies in a firm. The sample will include only mothers that have kids under 5 years.*
Snowball sampling	The population is hard to access so we use participants to share a call for participation to others. We utilize a social network effect to reach the targeted sample size.	*We want to research the way of conducting business in social supermarkets as a new phenomenon in food supply chain. As there is no valid register of such organizations, we make the first contact with several managers of social supermarkets in our area and then ask them to share their contact information on other managers with us. We repeat this action with other social supermarkets' managers until we reach a certain number of members in our sample.*

Source: Own elaboration.

When designing research, we have to try to avoid sample biases. A **sample bias** is a situation in which some members of population have a greater chance to become sample members than others. In such a situation, the possibility of generalization of results to targeted population is limited. Findings from samples that are biased can be generalized only to the part of population with the same characteristics as the sample.

A sample bias can occur both in probability and in non-probability sampling methods. However, in probability sampling, bias, by definition, is usually lower than in non-probability sampling.

There are some suggestions on how to reduce sampling bias:
- clearly identify your survey goals,
- clearly define characteristics of your targeted population,
- strictly define the requirements for your target sample members, do it in accordance with the characteristics of the population,
- ensure potential respondents an equal chance of taking part in your survey,
- regularly check if the profile of respondents fits your survey's goals,
- include more respondents (enlarge sample size),
- make online questionnaire as short as possible and accessible form different devices,
- send several follow ups to potential sample members to motivate them to respond.

Stratified sampling: example

We intend to conduct research on food waste by young people in six selected countries in Central and Eastern Europe: Czech Republic, Hungary, Poland, Slovak Republic, Bulgaria and Croatia. It was hypothesized that the respondents' answers would depend on their age and gender. Using the Formula (1), having the data presented in the table 4 and assuming that p is 0.5, the margin of error is 2% and the confidence level is 95%, the minimum sample size was calculated at the level of 2,400 respondents. The question is, how many people from each country (by age and gender) should be included in the research sample?

To answer this question, data on the population of interest to us (N = 14,685,442 people) was collected and differentiated by age and gender (Table 2.8).

Table 2.8. Population of young people in selected countries in 2018

	20–24		25–29		30–34	
	women	men	women	men	women	men
Czech Republic	249,465	261,933	328,088	344,934	350,984	373,507
Hungary	276,356	293,067	302,709	323,479	296,272	312,200
Poland	1,058,343	1,101,924	1,296,586	1,344,572	1,518,295	1,561,848
Slovak Republic	151,195	158,815	188,303	196,327	206,880	216,384
Bulgaria	153,262	163,210	209,886	222,545	232,002	247,635
Croatia	117,329	123,054	117,514	122,789	130,048	133,702

Source: Demography and Population.

The percentage structure of the population was then calculated as presented in Table 2.9.

Table 2.9. Percentage structure of young people in selected countries in 2018 (in %)

	20–24		25–29		30–34	
	women	men	women	men	women	men
Czech Republic	1.70	1.78	2.23	2.35	2.39	2.54
Hungary	1.88	2.00	2.06	2.20	2.02	2.13
Poland	7.21	7.50	8.83	9.16	10.34	10.64
Slovak Republic	1.03	1.08	1.28	1.34	1.41	1.47
Bulgaria	1.04	1.11	1.43	1.52	1.58	1.69
Croatia	0.80	0.84	0.80	0.84	0.89	0.91

Source: Own calculation.

In the next step, considering the determined structure of the population and the minimum size of the sample, the number of women and men to be tested in each age group were determined. The results are included in Table 2.10.

Table 2.10. Number of respondents needed to be tested in each subgroup

	20–24		25–29		30–34	
	women	men	women	men	women	men
Czech Republic	41	43	54	56	57	61
Hungary	45	48	49	53	48	51
Poland	173	180	212	220	248	255
Slovak Republic	25	26	31	32	34	35
Bulgaria	25	27	34	36	38	40
Croatia	19	20	19	20	21	22

Source: Own calculation.

Then we apply random sampling method or systematic sampling method to select a certain number of people from each subgroup we created.

2.4. Variables—first view

2.4.1. Introduction

Researchers usually start their analysis with investigating single variables. Such an initial analysis may answer many basic questions, e.g., how many observations fall in the categories, what is the structure of values, what are the typical values or ranges, or what is the dispersity of data.

The first view of the variable includes checking its frequency table. This works well if the variable does not have a large number of values. From the frequency

table we can easily read the minimum and maximum value, missing data, the pattern of frequencies and percentage.

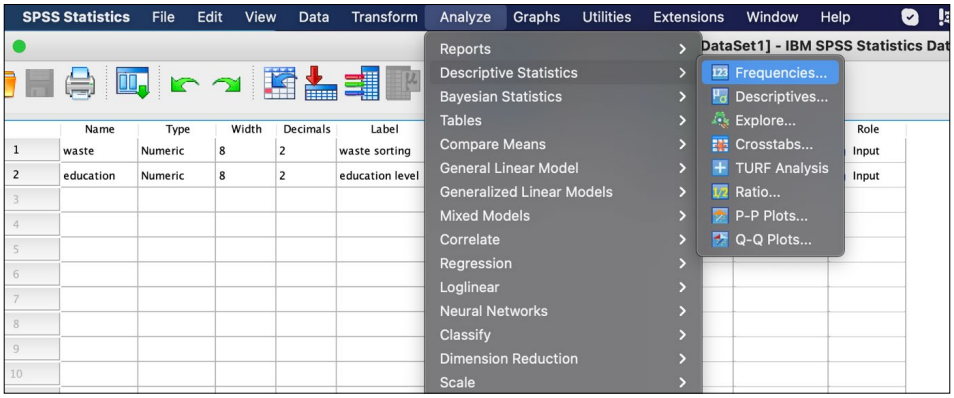

Figure 2.1. Screenshot of obtaining frequency table

Source: Own elaboration.

Frequencies

Statistics

waste sorting

N	Valid	160
	Missing	0

waste sorting

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 no, everything goes to general waste	49	30,6	30,6	30,6
	2.00 we sort waste into not more than two categories	78	48,8	48,8	79,4
	3.00 we sort waste into more than two categories	33	20,6	20,6	100,0
	Total	160	100,0	100,0	

Figure 2.2. Screenshot of frequency table (waste sorting in households)

Source: Own elaboration.

Within the same frequencies command, we can obtain descriptive statistics of the variable. It just requires clicking the button Statistics and choosing the statistics appropriate for the measurement level of the investigated variable.

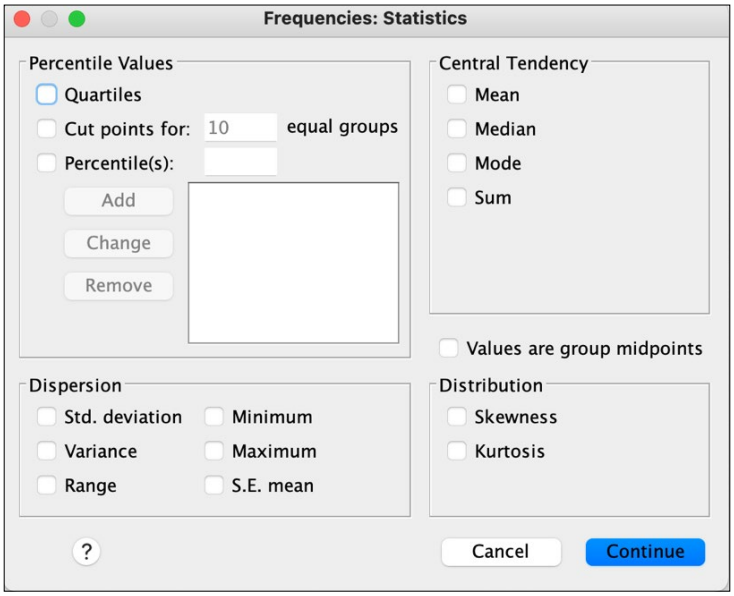

Figure 2.3. Screenshot of the dialog box of basic statistics within the Frequency command

Source: Own elaboration.

In Figure 2.4. and Figure 2.5. you can see the table including statistics for the ordinal and ratio level of measurement.

Statistics

Attitudes toward waste sorting

N	Valid	200
	Missing	0
Median		3,00
Mode		2
Percentiles	25	2,00
	50	3,00
	75	3,00

Figure 2.4. Descriptive statistics for ordinal variable (Frequency command)

Source: Own elaboration.

Statistics

sustainability

N	Valid	200
	Missing	0
Mean		6,33
Std. Error of Mean		,198
Median		6,00
Mode		6[a]
Std. Deviation		2,800
Variance		7,840
Skewness		,010
Std. Error of Skewness		,172
Kurtosis		-,162
Std. Error of Kurtosis		,342
Range		14
Minimum		0
Maximum		14
Sum		1266
Percentiles	25	4,25
	50	6,00
	75	8,00

a. Multiple modes exist. The smallest value is shown

Figure 2.5. Descriptive statistics for ratio variable (Frequency command)

Source: Own elaboration.

The frequency tables can be visualized with the chart. Same frequency procedure allows for producing charts (clicking the button charts). We can choose between:
- a pie chart,
- a bar chart,
- a histogram.

Pie charts are designed for nominal data. They can show the percentage of categories, but they are difficult to compare. They are not clear with the bigger number of categories. And even with the smaller number of categories we can be mistaken assessing the structure. So even for the nominal data the bar chart is really worth considering as the default one. To explain the issue, compare the structure in two pie charts presented below.

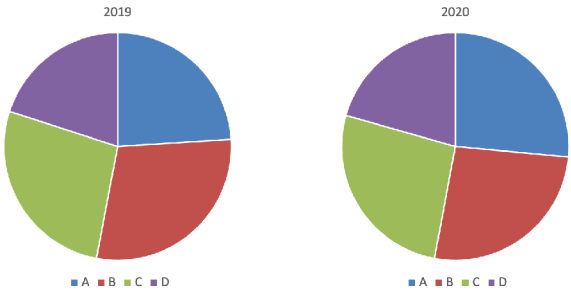

Figure 2.6. Pie charts

Source: Own elaboration.

Bar charts are designed for nominal and ordinal data. They show differences between frequencies of categories much better. The vertical axis can show counts or percentages. Compare the same data as presented above in the form of bar charts, the differences within the chart and between the charts are much clearer now.

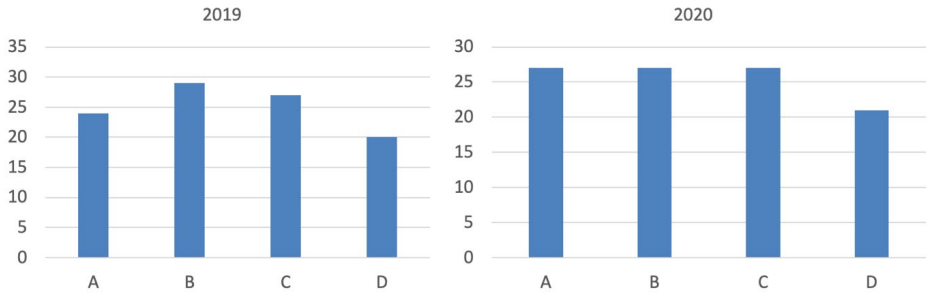

Figure 2.7. Bar charts

Source: Own elaboration.

The histograms are designed for interval and ratio levels of measurement. They are similar to bar charts, but bars are touching (continuous data), and every bar shows the frequency of observation within the interval.

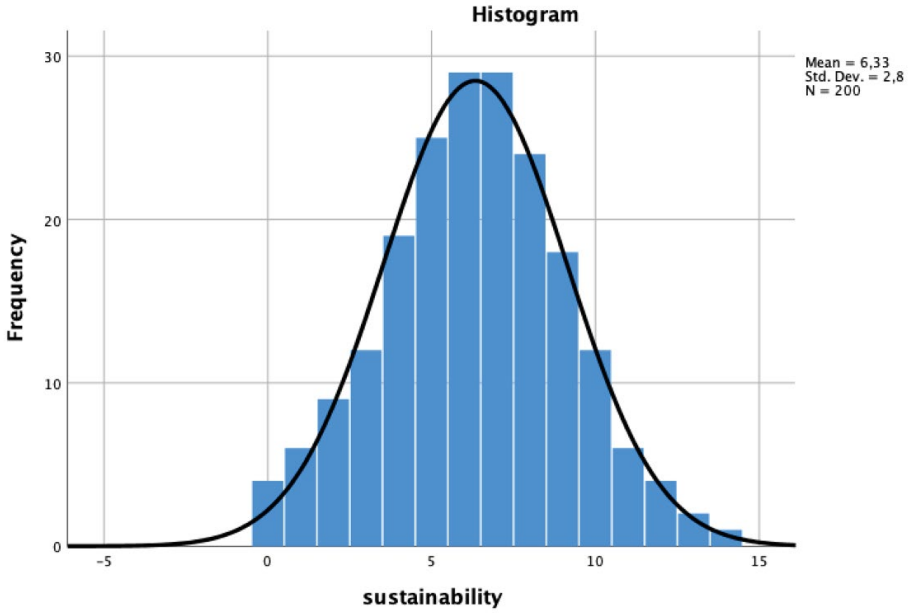

Figure 2.8. Histogram with the added normal curve

Source: Own elaboration.

2.4.2. Box and Whiskers chart

There is one more popular and very informative way to visualize variables. The chart is called box and whiskers chart, five numbers chart or box chart. What we need for constructing the chart is a minimum and a maximum, first, second and third quartile. The box is drawn having its bottom as the first quartile, the upper limit as the third quartile and the thick line within this range presenting the median.

The minimum and maximum are presented in the form of whiskers. It sounds very easy and can be hand-drawn, but SPSS definitely helps to produce it.

If the minimum (maximum) is considered an outlier, the whisker will present the smallest (highest) value not exceeding the 1.5 of interquartile range. And all the outliers will be shown as small circles (closer distance outliers) or asterisks (far distance outliers). You can produce the box and whiskers chart using the Chart menu in SPSS (both Charts / Legacy Dialog / Boxplot and Chart / Chart Builder / Boxplot lead to this), but a very handy way is available in Analyse / Descriptive Statistics / Explore. It produces a bunch of useful descriptive statistics, and two graphical summaries: Stem and Leaf (not described here), as well as Box and Whiskers chart. Just drop the investigated variable into the field of Dependent list. If you want to compare the results splitting your data (e.g., for females and males) you have to drop this variable into the factor list. There is one more field there, the Label Cases by. It tells SPSS how to name the outliers. If you leave this field empty, every outlier will get the number of rows from the data file, so you can spot it if needed. If you drop the variable there, the outliers will be named according to their value of the chosen variable.

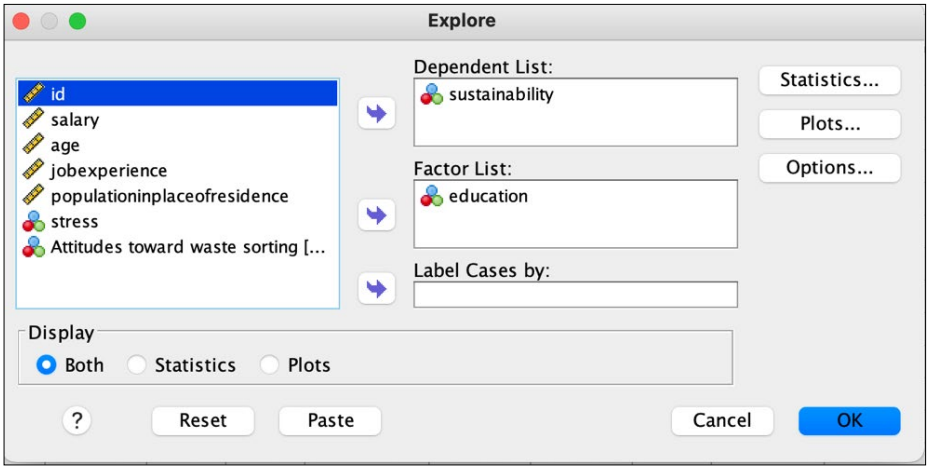

Figure 2.9. Dialog box for Explore (including Box and Whiskers chart)

Source: Own elaboration.

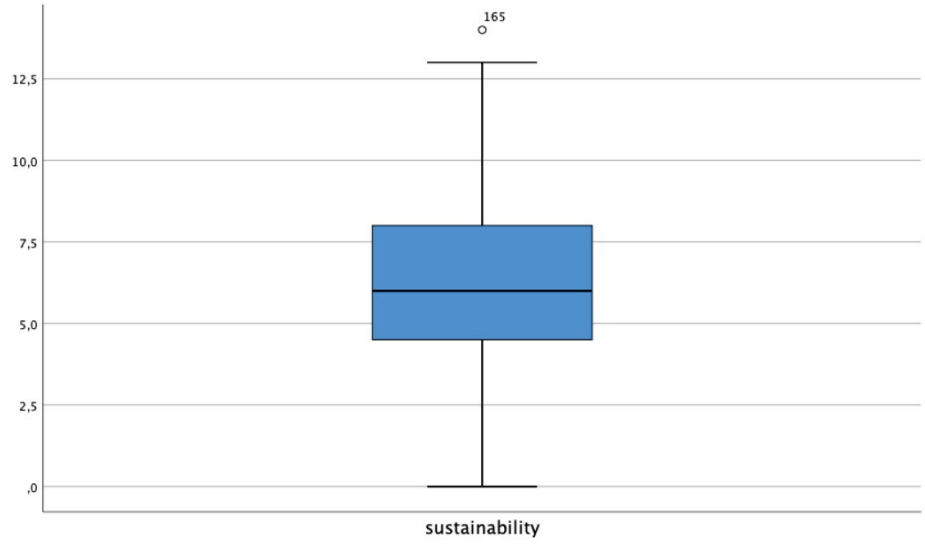

Figure 2.10. Box and Whiskers chart—single box

Source: Own elaboration.

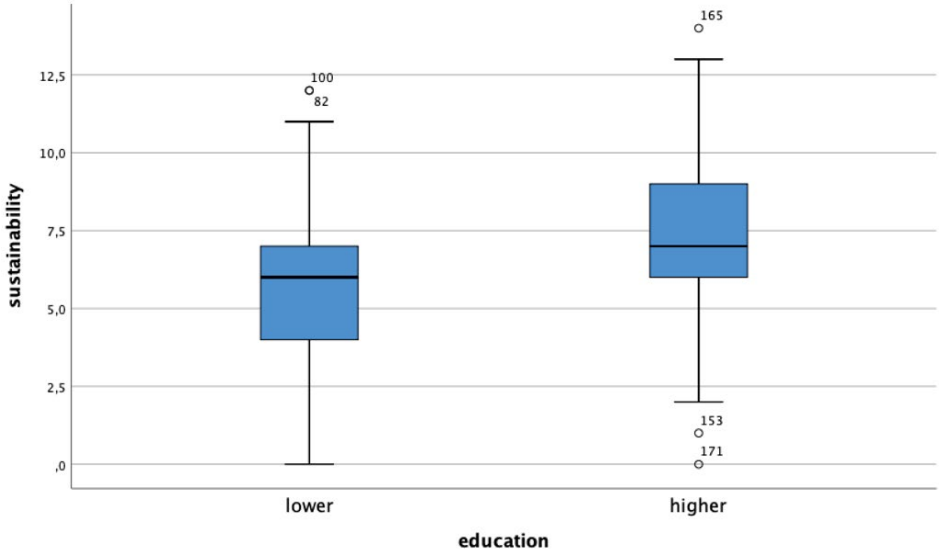

Figure 2.11. Box and Whiskers chart—comparison (more boxes)

Source: Own elaboration.

The comparison box shows that even if the difference between medians is small, the typical 50% of the people (inside the box) with higher education have better sustainability knowledge compared to the people with lower education levels. We can observe two outliers in the group with lower education (having much higher knowledge) and three outliers in the group with higher education (two much lower, one much higher).

2.4.3. Crosstabs: percentage

Tables are the most popular way to present data. But this popularity does not mean, that readers always understand them correctly. In this chapter we will focus on the tables presenting relationships between two categorical variables.

A simple table usually presents the dependent variable and all its values as rows. The independent variable constitutes columns. This allows to check the counts or percentage in the cell where row and column cross.

The basic table may be obtained in SPSS in Analyse / Descriptive Statistics / Crosstabs.

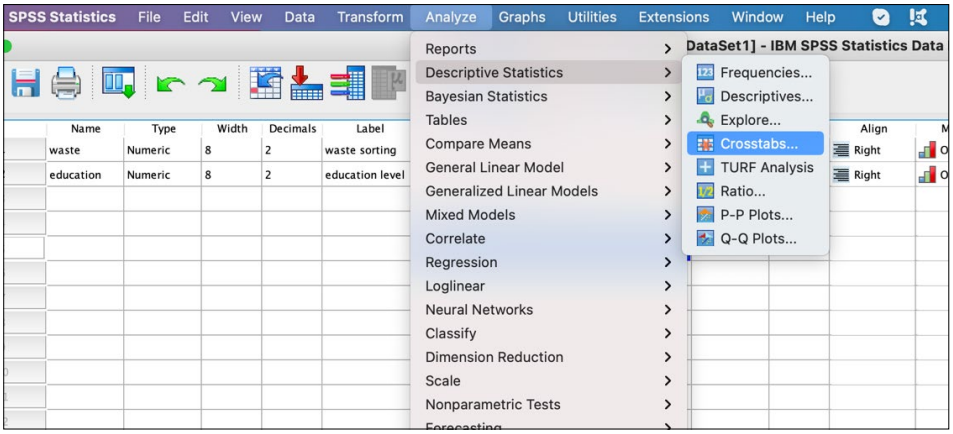

Figure 2.12. Obtaining the Crosstabs: path

Source: Own elaboration.

In the dialog box you have to drop the independent and dependent variable into fields called Row(s) and Column(s) as shown on the example in Figure 2.13.

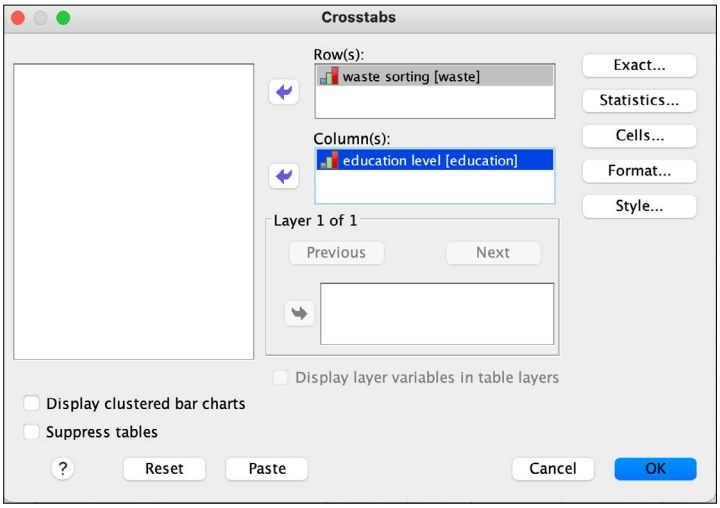

Figure 2.13. Dialog box of Crosstabs

Source: Own elaboration.

The default version produces a table with counts. Each cell of the table shows the number of observations meeting the condition of having a certain value of independent and certain value of dependent variable.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
waste sorting * education level	160	100,0%	0	0,0%	160	100,0%

waste sorting * education level Crosstabulation

Count

		education level			
		1.00 primary education	2.00 high school	3.00 university education	Total
waste sorting	1.00 no, everything goes to general waste	22	18	9	49
	2.00 we sort waste into not more than two categories	19	34	25	78
	3.00 we sort waste into more than two categories	10	12	11	33
Total		51	64	45	160

Figure 2.14. Crosstabulation with counts (default)

Source: Own elaboration.

We can read that 22 observations in the database meet both conditions: no sorting (all to general waste) and primary education. Below and on the right side of the table we can see the totals: the summarized observations from every row and from every column. We can read that we have 33 responses of sorting waste into more than two categories.

When presenting counts you can also display the expected values (the values if there are no dependencies between variables) or the differences between expected and observed values. Going further in this direction would lead us to chi-square, but it is not our intention.

Comparing counts may be difficult if the groups are not evenly divided. A much easier option is offered by presenting percentages. The table showing percentages are usually calculated as:

- row percentage,
- column percentage,
- table percentage.

The percentage direction in the table may be chosen in the dialog box within the procedure Crosstabs. This is under the button Cells. The default option is Observed, you can unclick it and choose the percentage.

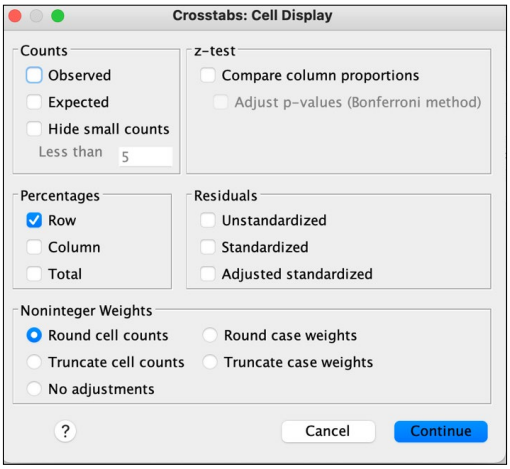

Figure 2.15. Dialog box of Cells options in the procedure of Crosstabs

Source: Own elaboration.

Row percentage tables take every row as the base for calculating percentages. It means that every row makes 100%. This shows the shares of all the values of the column variable in one single value of row variable. In Figure 2.16 it shows that 44.9% of all the people not sorting waste (everything goes to general waste) are people with primary education. As you can see every row sums up to 100%.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
waste sorting * education level	160	100,0%	0	0,0%	160	100,0%

waste sorting * education level Crosstabulation

% within waste sorting

		education level			
		1.00 primary education	2.00 high school	3.00 university education	Total
waste sorting	1.00 no, everything goes to general waste	44,9%	36,7%	18,4%	100,0%
	2.00 we sort waste into not more than two categories	24,4%	43,6%	32,1%	100,0%
	3.00 we sort waste into more than two categories	30,3%	36,4%	33,3%	100,0%
Total		31,9%	40,0%	28,1%	100,0%

Figure 2.16. Crosstabulation with row percent

Source: Own elaboration.

The tables displaying column percent are the most popular ones. They allow for comparing shares of row variable within the chosen value of column variable. This way we can easily read, in which group the waste management is more frequently accepted.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
waste sorting * education level	160	100,0%	0	0,0%	160	100,0%

waste sorting * education level Crosstabulation

% within education level

		education level			
		1.00 primary education	2.00 high school	3.00 university education	Total
waste sorting	1.00 no, everything goes to general waste	43,1%	28,1%	20,0%	30,6%
	2.00 we sort waste into not more than two categories	37,3%	53,1%	55,6%	48,8%
	3.00 we sort waste into more than two categories	19,6%	18,8%	24,4%	20,6%
Total		100,0%	100,0%	100,0%	100,0%

Figure 2.17. Crosstabulation with total percent

Source: Own elaboration.

The tables displaying table percent shows the percentage of observations meeting certain condition (having a given value of independent variable and given value of dependent variable) compared with the total number of observations in the table. Thus we have the 100% only if we sum up all the cells of the table.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
waste sorting * education level	160	100,0%	0	0,0%	160	100,0%

waste sorting * education level Crosstabulation

% of Total

		education level			
		1.00 primary education	2.00 high school	3.00 university education	Total
waste sorting	1.00 no, everything goes to general waste	13,8%	11,3%	5,6%	30,6%
	2.00 we sort waste into not more than two categories	11,9%	21,3%	15,6%	48,8%
	3.00 we sort waste into more than two categories	6,3%	7,5%	6,9%	20,6%
Total		31,9%	40,0%	28,1%	100,0%

Figure 18. Crosstabulation with table percent

Source: Own elaboration.

We can read from the table in Figure 2.18 that 13.8% of all observations meet both conditions: not sorting (all to general waste) and primary education. The totals below and on the right side of the table show the structure of column variable and row variable.

All the above tables (Figure 2.14, Figure 2.16, Figure 2.17, and Figure 2.18) were constructed from same database. Every table shows data according to certain expectations. Without knowing the direction of percenting the table one could be mistaken. Attention should be paid not only when constructing tables with percentages, but when reading someone's else results.

2.5. Visualization— Likert scale and some chosen charts

2.5.1. Visualization of the Likert scale

As mentioned in the chapter about questionnaire design, there are various scales helpful in collecting data. There are multiple ways to present data collected with these scales and showing them would probably lead to publishing a separate book. But at least the presentation of data from the Likert scale is worth mentioning as a highlighted example.

The Likert scale is a bunch of statements measured on the ordinal level. So, respecting the rules set in the previous chapter, the analysis should not go further than position measures. But the practice is different in this case. There is an assumption that the intervals within the scale are equal, and the mean and variance can be measured. This allows for various types of analyses, including the discriminant power of statements based on standard deviation or factor analysis based on variance.

This upgrade is often discussed and raises questions. I would suggest not upgrading the level of measurement if the scale uses less than five categories. The sample is small, and the scale uses verbal labels for all classes (not only for the extreme classes), or the distribution is heavily skewed.

There is one more often highlighted misunderstanding. The results from the Likert scale can be summed up and presented as mean or sum. This is the proper procedure and has nothing to do with the above-stated concerns.

Talking about the raw data collected using the Likert scale, the profile chart seems to be a proper tool for effective visualization. Furthermore, it can use both approaches: the ordinal and the interval assumption.

Such a chart displays the statements and the line with equally distributed points (units of the scale). For each statement, the mode or median (ordinal level of measurement approach) or mean (interval level of measurement approach) can be shown. Usually, all the points are connected by lines. This scheme allows for effective comparison of results in groups (e.g., results for females and males).

For better understanding, all the negative statements should be reversed, so every statement result on the right side of the chart means always agreeing more and having a stronger attitude.

The example below shows data about sustainable consumption behaviour collected using the questionnaire designed by Quoquab, Mohammad and Sukari (2019). There are two subgroups in the sample, the younger (red line) and the older (blue line) inhabitants of the camp. The lines are matching the means for every statement. It means, that the more to the right, the group agreed more with the statement. The gaps between lines show the differences between groups.

Figure 2.19. Profile chart from Likert scale

Source: Own elaboration.

2.5.2. Other examples of data visualization schemes

It is not our intention to decide which other charts are more practical, so three types of graphs are presented to show some options from the variety.

The scatterplot shows a relationship between two continuous variables. Every point on the scatterplot is one observation, and its X and Y measurements are the values of two variables.

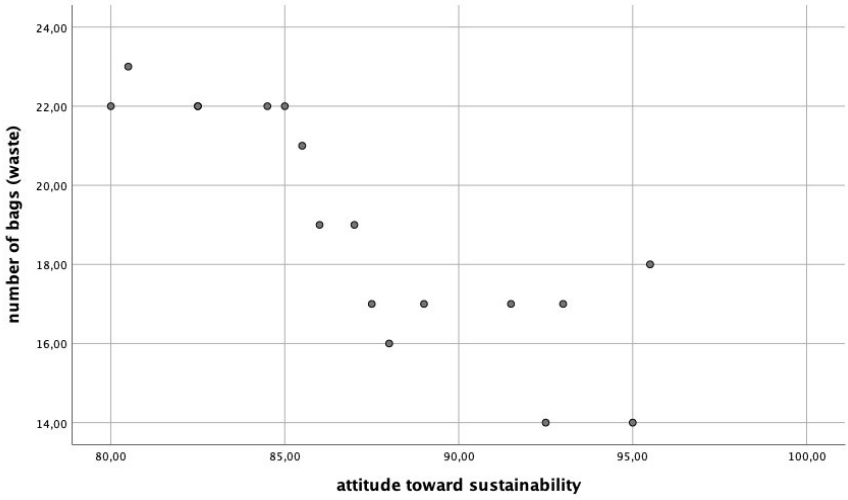

Figure 2.20. Scattterplot

Source: Own elaboration.

If the variables should suggest causation, the independent variable (cause) sets the horizontal axis, and the dependent variable (effect) sets the vertical axis. The point on the scatterplot can show additional information, through adding successive variables. Usually, additional information is brought by a difference of colour or shape of the points.

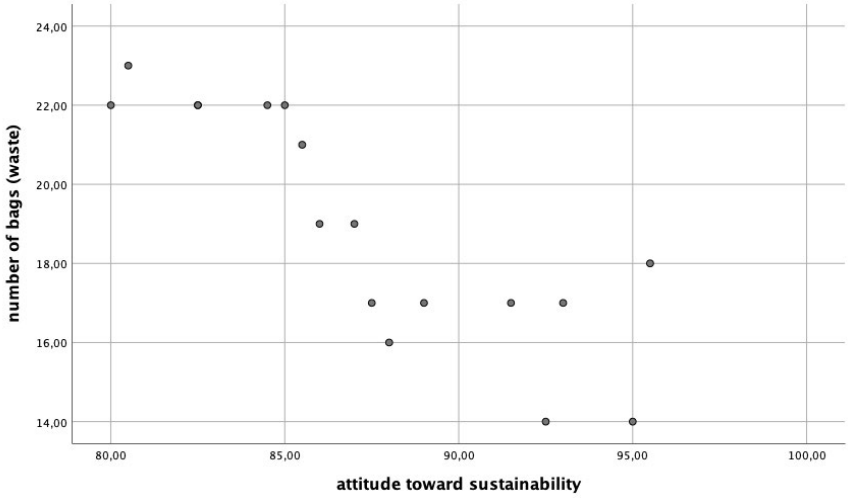

Figure 2.21. Scattterplot showing results for two groups

Source: Own elaboration.

Quite a popular extension of scatterplot is known as a bubble chart. In this case, an additional value (usually a continuous variable) is presented as the size of the point (bubble).

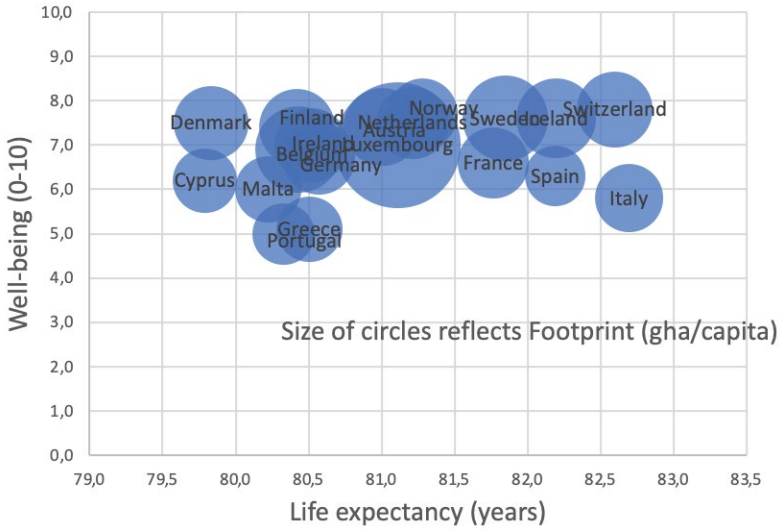

Figure 2.22. Scattterplot showing results for two groups

Source: Own calculations from Happy Planet Index data.

Questions / tasks

1. Which level of measurements gives more possibilities in analysis (more statistical techniques): ordinal or interval?
2. Which level of measurement are yearly CO_2 emissions in countries of the European Union (in tons) the examples of?
3. Can a median be applied for estimating the average emission?
4. Can mean and standard deviation be applied to describe the emissions in EU?
5. Give your own examples of variables measured on every level of measurement.
6. Considering the example presented in section 2.2.3, answer the following questions:
   - What scales were used in the questions?
   - Which questions and why may be irritating to consumers and will be skipped by them?
   - How can the questions be improved?
   - Create additional questions if necessary to verify the hypothesis.

- Create additional 3 questions that could be included in the part of characteristics.
- What methods can the data obtained in each of the questions be analysed with?

7. Face-to-face interview, telephone interviews, Web surveys or mail questionnaires—what types of research problems are they recommended for? Give some examples of methods.
8. Open-ended versus close-ended questions—what are the advantages and disadvantages of these questions? Give some examples of questions which can be used in the study on sustainable development.
9. How can the correctness of the questionnaire be further controlled before the pilot study?
10. How can you use the questions in the questionnaire to check if the respondent answers are reliable?
11. When designing the study on sharing economy, the researchers formulate, *inter alia*, the following research hypotheses:
    H1: The main reason to use Airbnb's services is to get to know the life of the inhabitants of a given country.
    H2: The consumer, having a choice of Airbnb and a hotel—assuming the same price—will choose Airbnb.
    H3: People who use Airbnb's services most often are people under 30, on low incomes, solo travellers.
    Prepare a fragment of the questionnaire that will verify the above hypotheses
12. Find an online questionnaire in scientific publications on selected aspects of sustainable development. Indicate what kind of questions and scales were used in that tool. Also make a substantive evaluation:
    - sharing economy,
    - ethical and unethical shopping behaviour of buyers,
    - sustainable development,
    - food waste,
    - shopping in second-hand shops,
    - renewable energy resources.
13. What topics are commonly threatening to respondents? How can a researcher ask about them?
14. In 2018, there were 640 third age universities in Poland, i.e., entities whose main goal is educational activity, integration and activation of older people in order to improve the quality of life and increase their participation in social life. A total of 113.2 thousand elderly people, including 95.4 thousand women, studied there. The greatest number of studying seniors are people aged 61–75 (71.9%), while people under 60 were 11.7%, and people aged 76 and over 16.4%. How

many retirees studying at such universities should be selected for the study on attitudes to the concept of lifelong learning assuming that the margin of error is 3% and confidence level 90%.

15. An average resident of a certain country in Europe saves an average of 6% of their salary. The average monthly salary at the beginning of 2021 was 1,450 euro. A certain financial institution intends to conduct a representative nationwide quantitative survey. It was hypothesized that the respondents' answers would depend on their average monthly savings.
    - Calculate how many respondents should be included in the sample of quantitative research, assuming a confidence level of 90% and the margin of error of 4% of the average savings. The statistical reports of banks also show that the average standard deviation of monthly savings in banks is 50 euro.
    - How will the sample size change if the confidence level is 95%?
    - How will the sample size change if the maximum acceptable error of estimate is reduced to 2%?
16. What sampling method should be used if the purpose of the quantitative study is to define forecasts for the sustainable development of air transport in selected countries of Central and Eastern Europe? The respondents in the survey are to be airport managers in the Czech Republic, Hungary, Poland, the Slovak Republic, Bulgaria, Croatia and Ukraine.
17. One international research agency intends to conduct quantitative research on the attitudes of people in several European countries towards the problems of sustainable development. It was assumed that the attitudes would depend on the level of education and age of the inhabitants. Select one country, find statistical data on a given society (in terms of education and age), then determine the minimum sample size and use stratified selection to determine the size of each sub-group.
18. What kinds of variables may be visualized using the bar chart?
19. Which chart would you suggest for the variable on the ratio level of measurement?
20. How can you interpret the boxes in the Box and Whiskers chart?
21. Explain the differences in three ways of presenting tables.
22. What characteristics of Likert scale should be considered before upgrading it from ordinal to the interval level of measurement?
23. In case of showing causation, how the independent and dependent variable should be presented in the scatterplot?
24. What is the difference between scatterplot and bubble chart?

References

Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it?. *Indian Journal Of Medical Specialities*, *4*(2), 330-333.

Aczel, A. D. (2009). *Complete business statistics* (7th ed.). New York: McGraw-Hill.

Barrow, M. (2017). *Statistics for economics, accounting and business studies*. Harlow: Pearson.

Bougie, R., & Sekaran, U. (2020). *Research methods for business. A skill building approach*. Hoboken: John Wiley & Sons.

Churchill, G. A., & Iacobucci, D. (2018). *Marketing research. Methodological foundations* (12th ed.). Nashville: Earlie Lite Book, Inc.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: Sage.

Dean, S., & Illowsky, B. (2013). *Introductory statistics*. Houston: Rice University, OpenStax College.

Demography and Population. Retrieved February, 2021 from https://stats.oecd.org/

Etikan, E., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1-4.

Happy Planet Index data. Retrieved May 19, 2021 from happyplanetindex.org

Kasiulevičius, V., Šapoka, R., & Filipavičiūtė, R. (2006). Sample size calculation in epidemiological studies. *Gerontologija*, *7*(4), 225-231.

Levy, P. S., & Lemeshow, S. (2008). *Sampling of populations: Methods and applications*. Hoboken: John Wiley & Sons.

McDaniel, C., & Gates, R. (2018). *Marketing research* (11th ed.). Hoboken: John Wiley & Sons.

Malhotra, N. K., & Birks, D. F. (2003). *Marketing research. An applied approach*. Harlow: Prentice Hall.

Neuman, W. L. (2014). *Social research methods: Qualitative and quantitative approaches*. Harlow: Pearson.

Quoquab, F., Mohammad, J., & Sukari, N. N. (2019). A multiple-item scale for measuring "sustainable consumption behaviour" construct: Development and psychometric evaluation. *Asia Pacific Journal of Marketing and Logistics*, *31*(4), 791-816.