

Advanced Quantitative Research Methods in Nursing





Faculty of Health Sciences

Advanced Quantitative Research Methods in Nursing

Authors

Lucija Gosak Leona Cilar Budler Roger Watson Gregor Štiglic

July 2024

Title	Advanced Quantitative Research Methods in Nursing
Authors	Lucija Gosak (University of Maribor, Faculty of Health Sciences)
	Leona Cilar Budler (University of Maribor, Faculty of Health Sciences)
	Roger Watson (University of Hull, Faculty of Health Sciences)
	Gregor Štiglic (University of Maribor, Faculty of Health Sciences)
Review	Graeme D. Smith (Saint Francis University)
	David R. Thompson (Queens University Belfast)
Language editor	Stojan Primožič (University of Maribor)
Technical editor	Jan Perša (University of Maribor, University Press)
Cover designer	Jan Perša (University of Maribor, University Press)
Graphic material	Sources are own unless otherwise noted. Gosak, Cilar Budler, Watson, Štiglic (authors), 2024
Cover graphics	Diversity-people-heads-group (7803661), author: geralt, pixabay.com, 2024
Published by Zaloźnik	University of Maribor University Press Slomškov trg 15, 2000 Maribor, Slovenija https://press.um.si, zalozba@um.si
Issued by Izdajatelj	University of Maribor Faculty of Health Sciences Žitna ulica 15, 2000 Maribor, Slovenia https://press.um.si, zalozba@um.si
Izdaja	First edition
Publication type	E-book
Available at	http://press.um.si/index.php/ump/catalog/book/899
Izdano	Maribor, Slovenia, July 2024



© University of Maribor, University Press

/ Univerza v Mariboru, Univerzitetna založba

Text © Gosak, Cilar Budler, Watson, Štiglic (authors), 2024 This work is licensed under the Creative Commons Attribution 4.0 International License.

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

Any third-party material in this book is published under the book's Creative Commons licence unless indicated otherwise in the credit line to the material. If you would like to reuse any third-party material not covered by the book's Creative Commons licence, you will need to obtain permission directly from the copyright holder.

https://creativecommons.org/licenses/by/4.0

```
CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor
616-083(0.034.2)
ADVANCED Quantitative Research Methods in Nursing [Elektronski vir] / authors
Lucija Gosak ... [et. al.]. - 1st ed. - E-knjiga. - Maribor : University of
Maribor, University Press, 2024
Način dostopa (URL): https://press.um.si/index.php/ump/catalog/book/899
ISBN 978-961-286-888-8 (Pdf)
doi: 10.18690/um.fzv.2.2024
COBISS.SI-ID 202111491
```

ISBN	978-961-286-888-8 (pdf)
DOI	https://doi.org/10.18690/um.fzv.2.2024
Cena Price	Free copy
For publisher Odgovorna oseba založnika	Prof. Dr. Zdravko Kačič, Rector of University of Maribor
Attribution Citiranje	Gosak, L., Cilar Budler, L., Watson, R., Štiglic, G. (2024). <i>Advanced Quantitative Research Methods in Nursing</i> . University of Maribor, University of Maribor Press. doi: 10.18690/um.fzv.2.2024



Table of Contents

1	Data and Quantitative Research Methods in Nursing	1
2		2
Z	Quantitative Study	
2.1		I I
2.2	Methodology and Methods	
2.3	Nesuits	12
2.4	Discussion	12
2.5	Conclusion	13
3	Sampling and Sample Size	15
3.1	Calculating the Required Sample Size Using SPSS	
		0.5
4	Review of Statistical Basics	
4.1	Start the IBM SPSS Software Package	
4.2	Creating New Variables	
4.3	Steps for Conducting Descriptive Statistics in SPSS	
4.4	Statistical Inference	
4.5	Performing Independent Samples t-1 est in SPSS	
4.6	Performing the Kruskal Wallis Test in SPSS	
4./	Performing the ANOVA Test in SPSS	
4.8	Performing the Correlation Test in SPSS	
4.9	Performing the Chi-Square Test in SPSS	
4.10	Performing the Mann-Whitney U Test in SPSS	
5	Regression	61
5.1	Linear Regression	62
5.2	Steps to Conduct the Linear Regression in SPSS	65
5.3	Logistic Regression	67
5.4	Steps to Conduct the Logistic Regression in SPSS	68
5.5	Multiple Regression	71
5.6	Steps to Conduct the Multiple Regression in SPSS	72
6	Questionnaire Validation	
61	Questionnaire Translation	77
6.2	Questionnaire Development	
6.3	Reliability vs. Validity	
0.5	remainly vo. valuely management	

ii	ADVANCED QUANTITATIVE RESEARCH MET	HODS IN NURSING
6.4	Test-Retest Reliability	
6.5	Inter-Rater Reliability	
6.6	Content Validity	
6.7	Construct Validity	
7	Reliability Analysis	85
7.1	Intraclass Correlations	
7.2	Inter-Rater Reliability	
7.3	Test-Retest Reliability	
7.4	Intra-Rater Reliability	
8	Factor Analysis	95
8.1	Steps to Conduct PCA in SPSS	
Refer	rences	



Symbols and Abbreviations Used

CI	confidence interval
CVI	content validity
H0	null hypothesis
Ha, H1	working hypothesis
I-CVI	item content validity index
Μ	Mean
Mo	Mode
Me	Median
Р	P-value (statistical significance)
RCT	Randomized controlled trial
S-CVI	scale content validity index
S-CVI/Ave	scale-level content validity index based on the average method
S-CVI/AU	scale-level content validity index based on the universal agreement
	method
SD	Standard deviation



1 Data and Quantitative Research Methods in Nursing

Nurses use and generate a lot of data in their daily work with patients. The main goal of patient care is to provide high-quality and evidence-based nursing care, so nurses are obliged to keep up to date with and apply the latest studies and evidence to their work. It is, therefore, important that nurses have the knowledge and understanding of data and data analysis techniques, as this is the only way to understand the latest evidence. Presented here is a short guide that includes basic information on quantitative research methods and information on statistical tests with which nurses need to be familiar in their work.

The data used in the book is based on data collected using the Self-Care of Diabetes Index (SCODI) questionnaire (Ausili, et al., 2017). The data were collected as part of a pilot study on patients with diabetes in Slovenia. The data are intended for educational and research purposes only.

There are many hands-on examples included to make the materials more engaging. We hope you will find this introduction to quantitative data analysis for nurses useful.



2 Quantitative Study

The following section describes the basic characteristics of quantitative research and breaks down the different types of quantitative research. The final part of the chapter gives a brief overview of how to produce a quantitative research report. The learning objectives of this chapter are to enable the reader to understand the basic concepts of quantitative research, to learn about the different types of quantitative research and to acquire the basic skills to produce a quantitative research report.

In the natural and social sciences, quantitative research is the systematic empirical study of observable phenomena using statistical, mathematical, or computational techniques. Figure 2.1 shows different aspects of the quantitative study.

The basic characteristics of the quantitative study (in comparison to qualitative) are:

- Data are collected using structured research instruments (surveys, structured interviews, experiments, observations, reviews of records or documents where any variable can be measured);
- it answers the questions: What? When?;
- it is mostly based on numbers;

- usually, a deductive process is used to test the predefined concepts, constructs, and hypotheses;
- the results are based on larger sample sizes that are representative of the population but provide less in-depth information;
- the research study can usually be repeated;
- the researcher has a clearly defined research question which demands objective answers;
- all aspects of the study are carefully designed before data collection;
- data are in the form of numbers, and results are often arranged in numerical tables, graphs, figures, or other non-textual form;
- broader understanding of concepts, predicting future results, or exploring a causal relationship;
- provides the observed effects (interpreted by the researchers);
- the researcher uses tools, such as questionnaires or computer software, to collect numerical data;
- usually, more time is needed in the planning phase and less in the analysis phase;
- the overall goal of a quantitative study is to classify characteristics, count them, and build statistical models to try to explain what we observe (Watson, 2015; Holton & Burnett, 2005; Reaves, 1992).



Figure 2.1: Different Aspects of Quantitative Study



The types of quantitative study are shown in Figure 2.2.

Figure 2.2: Quantitative Study Types

Researchers sometimes classify types of studies differently. Thus, instead of descriptive and correlation studies, they often refer to observational and quasiexperimental and experimental research intervention studies. Observational studies include cross-sectional studies, case studies with controls, cohort studies (retrospective, prospective), case studies, and case series studies. The experimental study is used to determine whether there is a cause-and-effect relationship between variables.

The following are basic definitions of the most common quantitative study types:

- Randomized controlled trial (RCT): In randomised controlled trials, participants are divided into an experimental group and one or more control groups. The only difference between the groups is the intervention being studied (Fig. 2.3). Studies can be known as single-, double- or triple-blind randomised controlled trials. Single-blind means that the participants do not know which group they are allocated to, double-blind means that the researchers collecting the data do not know which of the participants belongs to a control or intervention group and triple-blind means that the researchers analysing the data do not know which group anyone was in (Stanley, 2007; Spieth, et al., 2016).



Figure 2.3: Randomized Controlled Trial

Pre- and post-test study (without randomization): This is a type of quasiexperimental study where the outcome of interest is measured twice, with variables measured before and after the intervention (**Fig. 2.4**). Participants are selected in a non-random way (Dugard & Todman, 1995; Marsden & Torgerson, 2012).



Figure 2.4: Pre- and Post-Test Study

 Cross-sectional study: A cross-sectional study is a study that analyses data on variables collected over a given period (Fig. 2.5). It also measures the outcome and exposure of the study participants at the same time. Participants are selected based on inclusion and exclusion criteria established for the study.



Figure 2.5: Cross-Sectional Study

Longitudinal study: Data are collected over time at different time points (Fig. 2.6). The researchers observe and collect data without trying to influence the variables.



Figure 2.6: Longitudinal Study

- Cohort study: In cohort studies, researchers investigate the links between risk factors and outcomes.
- Retrospective cohort study: A retrospective cohort study is a longitudinal cohort study that compares cohorts of individuals where one group is exposed to a factor and the other is not. This is to determine the impact of that factor on the phenomenon. They have the advantage that they can be carried out immediately because they look at the situation retrospectively but may have less control over the exposure of the participants to the factor.
- Prospective cohort study: This is a longitudinal study that follows groups of similar individuals over time, differing in a trait/factor that may influence a particular condition.



Figure 2.7: Types of Studies According to the Time of Data Collection

 Case-control study: Comparison of two groups of people, where one group has the condition in question and the other group does not (Fig. 2.8). Here, researchers look at the history of the participants (health, lifestyle) to see which factors are associated with a particular condition.



Figure 2.8: Case-Control Study

- Correlation study: Correlational study is used to study the relationships between different subjects and variables. A correlational study is nonexperimental, meaning that none of the variables are influenced or controlled by the researcher. We distinguish between positive or negative correlation and no correlation. Correlational studies cannot prove cause-and-effect relationships between variables.
- Survey study: Descriptive surveys are used to formulate a hypothesis, but only after all the necessary data have been collected. Participants in the survey answer a set of questions asked by the researcher. These surveys are mainly carried out at the beginning of projects or larger research studies to identify the problem and the direction of the study.

Methods, in general, can be divided into data collection methods, data analysis methods, data synthesis methods, and data display methods. These depend on the study plan or type of study.

Quantitative methods emphasize the use of objective measurements and statistical, mathematical, or numerical analysis of data collected through surveys or questionnaires and or the manipulation of existing statistics using computational techniques. The quantitative study focuses on the collection of numerical data and their generalization by groups of people or the explanation of a particular phenomenon.

Before designing a quantitative survey, we need to decide on a study plan/design that will determine how the results are collected, analyzed, and interpreted. In a descriptive study, we follow the following rules: the object of study is usually measured once; the purpose is only to establish links between the variables; the study may include a sample population of hundreds or thousands of subjects to provide a valid estimate of the overall relationship between the variables. The experimental design includes subjects measured before and after a particular treatment. The sample population can be very small and purposefully selected, and it is designed to establish causality between variables. **Figure 2.9** shows the steps in a quantitative study.



Figure 2.9: Quantitative Study Process Source: (Kmetec, et al., 2019; Grove, et al., 2015)

The main features of a quantitative study concerning individual sections, when reported in a scientific paper, are described below.

2.1 Introduction

The introduction to a quantitative study is usually written in the present tense and from a third-party perspective. It covers the following data:

- Identifies the research problem as with any academic study, we should clearly and concisely state the research problem.
- Reviews the literature review studies on this topic, synthesize key topics, and, if necessary, list studies that use similar methods of investigation and analysis. Consider where there are key knowledge gaps and how your study helps fill those gaps or clarify existing knowledge.
- Describes the theoretical framework give an outline of the theory or hypothesis on which your study is based. If necessary, identify unknown or complex terms, concepts, or ideas and provide relevant background information to place the research problem in an appropriate context.

2.2 Methodology and Methods

The sketch on the quantitative study methods should describe how each goal of your study will be achieved. Provide sufficient detail to enable the reader to make an informed assessment of the methods used to obtain results related to the research topic. The section on methods should be presented in the past tense:

- Sample and sampling where the data comes from; note where gaps exist or what has been ruled out. Follow the procedures used to select them.
- Data collection describe the tools and methods used to collect information and identify variables; describe the methods used to obtain the data; note if the data already existed or you collected them yourself. If you collected them yourself, describe what instrument you used and why. Note that no data set is complete - describe any limitations in data collection methods.
- Data analysis describe the procedures for data processing and analysis. Where
 appropriate, describe the specific analysis instruments used to study each

research objective, including the mathematical techniques and the type of computer software (e.g., Jamovi, JASP, Orange, R, etc.) used to process the data.

2.3 Results

The study findings must be reported objectively and in a concise and accurate format. In quantitative studies, it is common to use graphs, tables, and other nontextual elements to help the reader understand the data. Make sure that the nontextual elements do not stand apart from the text but are used to supplement the general description of the results and to clarify the key tense.

 Statistical analysis - how did you analyse the data? What were the key findings? The findings should be presented in a logical sequence. Describe but do not explain these trends or negative results; save this for discussion. The results should be presented in the past period.

2.4 Discussion

The discussion should be analytical, logical, and comprehensive. The discussion should combine your findings in conjunction with the findings of the literature review and place them in the context of the theoretical framework on which the study is based. The debate needs to be presented at present:

- Interpretation of results re-emphasize the study problem and compare the findings with the research questions on which the research is based. Did you confirm the announced results, or did you refute them with the data?
- Description of trends, comparison of groups, or relationships between variables - describe any trends that have emerged from your analysis and explain any unforeseen and statistically significant findings.
- Discussion of consequences what is the meaning of your results? Highlight key findings based on general results and consider findings that you consider relevant. How do the results help to fill gaps in understanding the research problem?
- Limitations describe any limitations or unavoidable biases in your study and, if necessary, consider why these limitations did not hinder the effective interpretation of the results.

2.5 Conclusion

Completing studies means summarizing the main findings of your research and giving a final comment or guidance:

- **Summary of findings** do not report statistics; provide a narrative summary of key findings and synthesize answers to your research questions.
- Recommendations if appropriate for the aim of the research, link the key findings to policy recommendations or actions to be taken in practice.
- Future research consider the need for future research related to study limitations or other gaps in the literature that have not been addressed in your study.

Different types of quantitative studies will demand different structures and requirements for reporting the results. Therefore, it is suggested to follow EQUATOR's (https://www.equator-network.org) internationally accepted guidelines for the presentation of different quantitative studies.



3 Sampling and Sample Size

The following section presents the basic features of sampling, including the theoretical background and practical applications. The different types of sampling such as simple random sampling, stratified sampling, systematic sampling and other relevant methods are discussed in detail. It also explains the process of sample size calculation using SPSS software. The learning objectives of this chapter are to enable the reader to understand the basic concepts and importance of sampling in the research process. The reader will become familiar with the different types of sampling and their characteristics. It will also provide practical knowledge of how to perform sample size calculations using SPSS. It also aims to develop the ability to evaluate critically and select the appropriate sampling method according to specific research needs, which is crucial for conducting high quality and reliable research.

The statistical population consists of all the participants in the group about which you want to draw inferences by observation. The statistical sample represents the individual part of the group (subgroup) from which the data will be collected (**Fig. 3.1**). The statistical unit represents one individual member of the population (element of the set).



Figure 3.1: Statistical Sample

Sampling methods fall into two broad groups: probability samples, where each element has a known probability of occurring in the class, and non-probability sampling, where the above condition is not met.

Simple random sampling: Individuals are randomly selected from the total population. This gives a sample that is not necessarily representative (Fig. 3.2) (Simkus, 2022).



Population

Figure 3.2: Simple Random Sampling

 Stratified sampling: The population is divided into classes according to certain characteristics (e.g., sex, age class, etc.). Then, random sampling (proportional or disproportionate) is carried out from each class.



Figure 3.3: Stratified Sampling

 Systematic sampling: The first element is selected randomly, followed by every kth element (k = population size/sample size).



Population

Figure 3.4: Systematic Sampling

 Cluster sampling: Use known information on the grouping of elements (sampling by provinces, localities, communities, by schools, classes, etc.).



Figure 3.5: Cluster Sampling

 Snowball sampling: First, we select a smaller sample of more accessible individuals, who are asked to disseminate the sample or questionnaire through their contacts.



Figure 3.6: Snowball Sample

- **Convenience sampling:** A convenience/occasional sample simply includes individuals who are most accessible to the researcher (Simkus, 2022).



Figure 3.7: Convenience Sampling

 Purposive sampling: the researcher uses his/her expertise to select the sample that is most useful for the purposes of the research.



Figure 3.8: Purposive Sampling

3.1 Calculating the Required Sample Size Using SPSS

One of the frequently asked questions during the study planning phase is, "How many participants do I need to test my hypotheses?". The answer to this question is not straightforward. In fact, we could write a whole book on different techniques of so-called power analysis that can be used to estimate the minimal required sample size for a given hypothesis. Calculating the sample size is key to ensuring accurate and reliable results, as it reduces the error rate and increases the precision of the estimates (Andrade, 2020).

As already mentioned in the initial chapters, to test the hypothesis, we need to know some characteristics of our data – especially the type of variables and distribution of collected data for the variables that are involved in the hypothesis testing.

There are many specialized computer programs available that allow us to conduct Power analysis. However, from IBM SPSS version 27, it is also possible to use the Power analysis function directly in the SPSS (*Analyze -> Power Analysis*).

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	Analyze Graphs Utilities	Extensions	Window	Help				
— =		10 1	Power Analysis	>	Means	>	0ne-Sample	T Test		
			Meta Analysis	>	Proportions	>	Paired-Samp	les T Test	_	
	A	0	Reports	>	Correlations	>	Independent-	Samples T Test		
		🂑 Q2	Descriptive Statistics	>	Regression	>		IOVA	DDI_C	SCODI_0
1	69	2	Bayesian Statistics	>	4		2 One-way Al	1014	4	
2	45	2	Tables	>	4		3 2	1	5	
3	59	1	Compare Means	>	-1		1 1	1	5	1
4	70	2	General Linear Model	>	4		2 1	1	-3	
5	43	2	Generalized Linear Medels	,	4		1 1	1	3	:
6	83	2		ĺ.	3		3 3	2	3	
7	45	2	WIXed Wodels	,	-1		2 1	-1	-3	-
8	54	1	Correlate	>	4		3 1	2	-3	-
9	46	2	Regression	>	1		2 1	1	5	
10	-1	1	Loglinear	>	1		2 1	2	3	
11	54	1	Neural Networks	>	4		3 1	1	5	1
12	46	1	Classify	>	1		2 1	1	3	
13	67	1	Dimension Reduction	>	1		2 2	1	3	
14	-1	2	Seele		1		3 3	1	5	
15	81	2	Scale	ĺ.	4		2 1	1	5	
16	51	2	Nonparametric Tests	,	1		3 2	1	1	
17	65	2	Forecasting	>	1		3 3	1	3	
18	64	2	Survival	>	4		2 1	1	5	1
19	-1	-1	Multiple Response	>	4		2 1	1	4	
20	59	2	Missing Value Analysis		4		4 3	2	1	
21	53	1	Multiple Imputation	>	-1		4 1	1	5	
22	49	1	Complex Samples		2		2 1	1	3	
Data View	Variable View		Simulation	,		-				

Figure 3.9: Power Analysis in SPSS

As it can be seen from Fig. 89, as of the IBM SPSS 28, the Power Analysis function in SPSS allows estimation of minimal sample size for four different groups of statistical tests:

- Comparison of means (univariate and bivariate t-tests, One-way ANOVA).
- Proportions (univariate and bivariate binomial tests).
- Correlations (Pearson, Spearman, partial).
- Regression (univariate linear).

Here, we provide a step-by-step demonstration of using the SPSS Power Analysis function to estimate the minimal required sample size for an independent t-test scenario.

After the selection of the *Analyze -> Power Analysis -> Means -> Independent-Samples t-Test*, we are asked to enter information for both compared groups (Fig. 3.10).

Power Analysis: Independent-Sample Means	×
Estimate sample size	P <u>l</u> ot
Single power value:	
O Grid gower values: Grid	
Grid values: None selected	
Group size ratio: 1	
O Estimate power	
Sample size for group 1: and group 2:	
Population mean difference:	
O Population mean for group 1: and group 2:	
Population standard deviations are	
Egual for two groups	
Pooled standard deviation:	
O Not equal for two groups Standard deviation for group 1:	
Test Direction	
Ondirectional (two-sided) analysis One of the side of th	
Significance level: 0.05	
OK <u>P</u> aste <u>R</u> eset Cancel Help	

Figure 3.10: Power Analysis for Independent-Samples t-Test

In this step, we will have to provide the "Single power value", which is usually set at 0.90 or 0.95. Figure 3.11 shows an example of the minimal sample size estimation for a scenario where we performed a pilot study on a small sample of students, measuring student satisfaction (measured on a scale from 1 to 5) of the undergraduate (group A, n=25) and postgraduate (group B, n=23) students. We obtained the following results:

- Average satisfaction level in group A was 4.5, and 4.8 in group B;
- Standard deviations of 0.9 (group A) and 0.8 (group B) were calculated;
- In the hypothesis we were interested only whether there are any differences in average satisfaction between undergraduate and post-graduate students (Nondirectional, two-sided analysis).

Power Analysis: Independent-Sample Means	×
Test Assumptions	Plot
● <u>E</u> stimate sample size	1 101
• Single power value: 0.95	
O Grid gower values: Grid	
Grid values: None selected	
Group size ratio: 1	
O Estimate power	
Sample size for group 1: and group 2:	
O Population mean difference:	
Population mean for group 1: 4.5 and group 2: 4.8	
Population standard deviations are	
\bigcirc E <u>q</u> ual for two groups	
Pooled standard deviation:	
Not equal for two groups	
Standard deviation for group 1: 0.9 and group 2: 0.8	
Test Direction	
Mondirectional (two-sided) analysis ■	
O <u>D</u> irectional (one-sided) analysis	
Significance level: 0.05	
OK Paste Reset Cancel Help	

Figure 3.11: Power Analysis Data Entry Example

Power Analysis - Independent Sample Means

Power Analysis Table								
				Test Assumptions				
	N1	N2	Actual Power ^b	Power	Std. Dev1	Std. Dev2	Mean Difference	Sig.
Test for Mean Difference ^a	211	211	.951	.95	.9	.8	.300	.05
a. Two-sided test.								

b. Based on noncentral t-distribution.

Figure 3.12: Results from the Power Analysis for an Independent Samples t-Test



4 Review of Statistical Basics

The next section introduces basic statistical concepts, such as hypotheses and variables, and provides basic information about basic descriptive statistics in SPSS software. The chapter also introduces the different statistical tests that are performed in SPSS. The learning objectives of this chapter are to enable the reader to understand basic statistical concepts and their importance in the research process. The reader will become familiar with formulating and testing hypotheses and will learn about the different types of variables and their influence on research results. They will also acquire basic skills in descriptive statistics in SPSS, including mean, median and standard deviation. Finally, the reader will learn how to perform and interpret various statistical tests in SPSS, such as t-tests, ANOVA, correlation tests and others. This chapter will provide the reader with a comprehensive understanding of key statistical concepts and methods.

Statistics is a fundamental tool in nursing, enabling nurses to analyse data systematically, which is key to improving the quality of patient care and making informed decisions. Understanding statistical concepts is essential for critically evaluating research studies that help to identify the most effective therapeutic approaches, establish associations between different risk factors and treatment outcomes, and optimise the management of healthcare resources. Understanding statistics also enables nurses to evaluate critically and apply research findings in their daily practice. This enables them to make more informed, evidence-based decisions that improve patient care and enhance the professionalism of their work:

- **Statistical population** These are all the objects (or "entities") that we observe; mathematically, the universal set.
- Statistical unit Individual member of the population; the element of the crowd.
- Statistical sample Part of the population; a subset.
- Statistical variable Property of individual members of the population.
- Statistical parameter Measured quantity describing a statistical population.

Hypothesis

- A hypothesis (plural hypotheses) is a proposed explanation for a phenomenon.
- An assumption that can be empirically verified.
- Consists of variables (descriptions, unit properties).
- Working, null, alternative, non-directional, directional.

We talk about a hypothesis when the problem is specific and well-defined and, therefore, can be formulated into a claim, which is then accepted or rejected by certain statistical tests. A hypothesis is an assumed statement of the relationship between two or more variables.

The null hypothesis (H_0) states that there is no statistically significant difference or relationship among variables. The working hypothesis (H_a, H_1) states exactly the opposite.

Using the p-value, we determine the degree of confidence with which we can reject the null hypothesis:

- **Simple:** shows a relationship between one dependent and one independent variable.
- Complex: shows the relationship between two or more dependent variables and two or more independent variables.
- Directional: the hypothesis assumes a relationship between two variables and is based on existing theory (E.g. Women are more confident in taking self-care for their diabetes).
- Non-directional: is used when no theory is involved. It is a statement that a relationship exists between two variables without predicting the exact nature (direction) of the relationship (E.g. There is a difference in diabetes self-care confidence between women and men).
- Null: provides the statement which is contrary to the hypothesis. It is a negative statement, with no relationship between independent and dependent variables (E.g. The average diabetes self-care confidence score is 60).
- Associative and causal: Associative hypothesis occurs when there is a change in one variable, resulting in a change in the other variable. Causal hypotheses propose a cause-and-effect interaction between two or more variables.

Variables

- To test hypotheses, we need to collect enough data for the variables involved in a hypothesis. A variable that causes a condition to develop is known as an independent variable because its values are independent of other variables. A variable that is influenced by an independent variable is called a dependent variable because its value depends on the cause (Fig. 4.1).



Figure 4.1: Type of Statistical Variables Source: (Liguori & Moreira, 2018)

4.1 Start the IBM SPSS Software Package

To run SPSS, double-click on the icon $\sum_{i=1}^{\alpha}$.

ew Files:				What's New:		
New Dataset New Database Query					Relationship Map	Variabie drigheit depter & Cender & Nartal roku
estore Points:				17.11	ton school	Category Count
Auto-Recovery Sep 14, 2021 10:00:10 AM	Data Editor	Untitled1*	* ^	Separated		ologe 0 1.000 - 800 - 0
	Output	Output1*		Widowed		Count
Auto-Recovery Sep 13, 2021 5:00:24 PM	Data Editor	sleep.sav	t	Divorced 4		
	Output	Output1*		Mar	rind Make Never framied NA	- 41 - 101 - 419 - 41 - 141 - 464 - 40 - 144 - 464 - 41 - 464
Auto-Recovery Sep 13, 2021 4:09:05 PM	Data Editor	sleep.sav	Ť	Relationshi to each othe	ip maps are useful for determini er by providing a visual represer	ing how variables relate ntation of the
Auto-Recovery Sep 13, 2021 1:31:45 PM	Data Editor	sleep.sav*	Ť	connections other.	s and influences that each node	and link has over each
	Output	Output2*				SPSS
Auto-Recovery Sep 13, 2021 1:21:45 PM	Data Editor	sleep.sav*	1	Help & Sup	pport <u>Tutorials</u>	Community
estore Points Recent Files	Sample Files			Visit SPSS Stat	tistics page to explore available	Give Feedback
			0	packages and	offers.	Report Issue

Figure 4.2: Start the SPSS Package

For SPSS analysis, we can use:

- data prepared for analysis in SPSS format (.sav file);
- data downloaded from other programs (e.g. Excel);
- data entered manually in the Data Editor window.

e Edit View Data Iransfo	orm <u>A</u> nalyze	Graphs Ut	ilities Extensio	ns <u>W</u> indow	Help		_		
New	,	2 =			। 🖬 🖣 🖓	۱ 💽 🌔	Q I		
Open Isseet Data		Data	el	Values	Missing	Columns	Align	Measure	Role
Import Qata		🛨 Internet D)ata						
General Open		Syntax							
Conen Restore Point		Output							
	041/54	Workboo	k						
Conse Conse	CHALO	Script							
Save .	CIII+5								
Save As									
Save All Data									
Save Restore Point									
Export	>								
Mark File Read Only									
Revert to Saved File									
File Enabled for Auto-Recovery									
E Collect Variable Information									
Rename Dataset									
Display Data File Information	>								
Gache Data									
Stop Processor	Ctrl+Period								
Switch Server									
Repository	>								
Print Preview									
A Drint	Ctriap								
Webser Dates									
weicome Dialog									
Recently Used Data	,								
Recently Used Eiles	,								

Figure 4.3: Importing Data Into SPSS

The Data View shows the dataset (Fig. 4.4). The columns represent variables, and the rows represent individual cases. Variable View shows information about the variables.



Figure 4.4: Data View Window in SPSS

Variable View includes the following information:

- name of variable;
- type of variable (e.g. string, numeric, etc.);
- number of signs (Width);
- number of decimal places (Decimals);
- variable brief description (Label);
- description of the variable value for coded categorical variables (Values);
- missing values (Missing);
- width of column (Columns);
- alignment of the variable values in Data View can be left, right or center justified (Align);
- type according to measurement (Measure; e.g., nominal, ordinal, or scale);
- role (e.g., input, target, both, etc.).

<u>F</u> ile	Edit	View Data	a <u>T</u> ransform	Analyze	Graphs	Utilities Extensions	Window Help					
	H			1	▙≡	P # =	🗮 📲 🎱 💽 🔍					
		Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1		Q1	Numeric	3	0	Age (years):	{-99, I don't know}	-991	8	🚟 Right	I Scale	S Input
2		Q2	Numeric	3	0	Gender:	{-99, I don't know}	-991	8	🖷 Right	🚓 Nominal	S Input
3		Q3	Numeric	3	1	Education:	{-99.0, I don't know}	-99.01.0	8	a Right	🚓 Nominal	S Input
4		Q4	Numeric	3	0	Are you employed?	{-99, I don't know}	-991	8	🗃 Right	🚓 Nominal	S Input
5		Q5	Numeric	3	0	What is your marital	{-99, I don't know}	-991	8	🔳 Right	🚓 Nominal	S Input
6		Q9	Numeric	3	0	Does anyone in you	{-99, I don't know}	-991	8	🚟 Right	🚓 Nominal	S Input
7		Q10	Numeric	3	0	How would you rate	{-99, I don't know}	-991	8	🚟 Right	🙈 Nominal	S Input
8		Q11	Numeric	3	0	How would you rate	{-99, I don't know}	-991	8	🚟 Right	🚓 Nominal	S Input
9		Q14	Numeric	3	0	Do you think you ha	{-99, I don't know}	-991	8	🚟 Right	\delta Nominal	S Input
10)	SCODI_C1	Numeric	3	0	Check your blood s	{-99, I don't know}	-991	8	🗃 Right	🙈 Nominal	🔪 Input
11		SCODI_C2	Numeric	3	0	When you have abn	{-99, I don't know}	-991	8	Right	🚓 Nominal	S Input

Figure 4.5: Variable View

4.2 Creating New Variables

a) Compute Variable

In the following section, we were interested in how are confident patients with diabetes in carrying out self-care activities and maintaining their health. Patients were asked to answer eleven questions using a Likert scale from 1 to 4. The formula for calculating the total score is given below:

Self-care Confidences = (SUM(SCODI_D1+ SCODI_D2+ SCODI_D3+ SCODI_D4+ SCODI_D5+ SCODI_D6+ SCODI_D7+ SCODI_D8+ SCODI_D9+ SCODI_D10+ SCODI_D11)-11)*(100/44)

To calculate a new variable from the given variables, we can use the Compute Variable function (Fig. 4.6).

ile <u>E</u> dit	View Dat	9	Transform	Analyze	Graphs	Utilities	Exte	Compute Variable				×
			Compute	e Variable								
	Mama		E Program	nmability T	ransformati	on		Target Variable:		Numeric Expression:		
1	O1	Nu	Count V	alues with	in Cases			Confidences =		((SCODI_D1 + SCODI_D2 + SCODI_D3 +	SCODI_D4	+ SCODI_D5 + SCODI_D6 + SCODI
2	02	Nu	Shift Val	lues						_D7 + SCODI_D8 + SCODI_D9 + SCODI_I	010 + SCO	DI_D11) - 11) * (100/44)
3	Q3	Nu		inte Com	Mariables			Type & Label				
4	Q4	Nu	Recode	into Same	variables			🖋 Age (years): [Q1] 🛆 🚄				
5	Q5	Nu	Recode	into Differ	ent Variable	\$		Sender: [Q2]				
6	Q9	Nu	Automat	tic Recode				& Education: [Q3]				Function group:
7	Q10	Nu	🛨 Create D	Dummy Va	riables			Are you employed		+ < > 7 8 9		All
8	Q11	Nu	📔 Visual 🗄	inning				& What is your mari				CDE & Negeostral CDE
9	Q14	Nu	Continual	Binning				Does anyone in y		- <= >= 4 5 6		Conversion
10	SCODI_C1	Nu	Prenare	Data for I	Andelina		,	A How would you rat		* = ~= 1 2 3		Current Date/Time
11	SCODI_C2	Nu	Report of		in a second		-	A How would you rat				Date Arithmetic
12	SCODI_C3	Nu	M Kank Ca	ases			_	Do you think you		/ & 0 .		Date Creation
13	SCODI_C4	NU	Date and	d Time Wi	zard			Check your blood				Functione and Special Variables
14	SCODI_CS	Nu	🔤 Create T	Ti <u>m</u> e Serie	S			& When you have a		** ~ () Delete	†	Eurodona and Opecial Valiablea.
16	SCODI_C0	Nu	📲 Replace	Missing)	alues			& When you have a				
17	SCODI_C8	Nu	🝘 Random	Number 9	Senerators			& When you have s			^	
18	SCODI C9	Nu	Run Per	uding Tran	forms		6	& If you find out that				
19	SCODI_C10	Nu	menc .	3	U	IT you Ting	out that	& If you find out that				
20	SCODI_D1	Nu	meric 3	3	0	Prevent hi	h or lov	After taking action				
21	SCODI_D2	Nu	meric 3	3	0	Follow adv	ice abo	A If you find out that			~	
22	SCODI_D3	Nu	meric 3	3	0	Take your	medicir	- , ,				
23	SCODI_D4	Nu	meric 3	3	0	Persist in	ollowing	(optional case selection cond	ditio	(ne		
24	SCODI_D5	Nu	meric	3	0	Monitor yo	ur bloor	li''' cohanan anda anan		,		
Data View	Variable Vie	~								OK Paste Reset Cancel	Help	

Figure 4.6: Compute New Variable

b) Record into Same or into Different Variables

If we want to map the existing values of a variable into new values, but do not want to create a new variable, we can use the Recode into Same Variables function. If we want to convert existing variables into new ones, we can use the Recode into Different Variables function (**Fig. 4.7**). In case we also have missing values, we need to add the conversion of the missing values back into the missing values.

Eile	Edit	<u>V</u> iew D	ata	Trar	nsform	Analyze	<u>G</u> raphs	Utilities	Exter	nsi	ons	Window
					Compute	e Variable						
_			-	+	Program	mability Tra	Insformatio	n				
	4	Name	N	2	Count V	alues within	Cases					alues don't
	2	02	N								1.00 1	don't
	2	03	N	_	Sni <u>π</u> va	ues					(-99, I	L don'
	4	04	N	X**X	Recode	into <u>S</u> ame	Variables				1-99.0	don't
	5	05	N	**)	<u>R</u> ecode	into Differer	t Variables				1-99 I	don't
	5	09	N	18	Automat	ic Recode					{-99	don't
-	。 7	Q10	N		Create F)ummy Vari	ables				{-99.1	don't
1	8	Q11	N		Viewel	inning					{-99, I	don't
	9	Q14	N	L.a	visual <u>L</u>						{-99, I	don't
1	0	SCODI_C1	N	K	Opt <u>i</u> mal	Binning					{-99, I	don't
1	1	SCODI_C2	N		Prepare	Data for Mo	odeling		>		{-99, I	don't
1	2	SCODI_C3	N	M	Ran <u>k</u> Ca	ises					{-99, I	don't
1	3	SCODI_C4	N		Date and	d Time Wiza	ard				{-99, I	don't
1	4	SCODI_C5	N		Create T	ime Series					{-99, I	don't
1	5	SCODI_C6	N			Mineter Market					{-99, I	don't
1	6	SCODI_C7	N	48	керіасе	wissing va	llues				{-99, I	don't
1	7	SCODI_C8	N	۳	Random	Number G	enerators				{-99, I	don't
1	8	SCODI_C9	N		Run Per	iding Transf	orms	Ctrl+	G		{-99, I	don't
1	9	SCODI_C1	0 N	ume	nc	3	U	IT you tina	OUT T.		{-99, I	don't
2	0	SCODI_D1	N	ume	ric	3	0	Prevent hi	gh or		{-99, I	don't
2	1	SCODI_D2	N	ume	ric	3	0	Follow ad	vice a		{-99, I	don't
2	2	SCODI_D3	N	ume	ric	3	0	Take your	medi		{-99, I	don't
2	3	SCODI_D4	N	ume	ric	3	0	Persist in	follow		{-99, I	don't
2	4	SCODI_D5	N	ume	ric	3	0	Monitor y	our bl.		{-99, I	don't
		< .		-								
Data	View	Variable	View									

Recode into Different Variables

		Numeric Variable -> Output Variable:	Output Variable
 n you mito out th Do you take ins If you find out th Prevent high or I Follow advice a Take your medi Monitor your blo Understand if y Recognize the Persist in monit Take action to a 	*	Confidences> Confidences_descri	Name: Confidences_descriptive Label: Change
Realuate if your			
Rersist in carryi 🗸		If (optional case selection condition)	
	OK	Paste Reset Cancel Help	

Figure 4.7: Recode Into a New Variable

×

We divided the values on a scale from 0 to 100 into four parts. 0 to 25 was low selfconfidence, 25 to 50 was insufficient self-confidence, 50 to 75 was adequate selfconfidence and 75 to 100 was good self-confidence.

Old Value	New Value
O <u>V</u> alue:	Value: System-missing
O <u>S</u> ystem-missing	O Cogy old value(s)
O System- or user-missing O Range: through C Range, LOWEST through value:	Olg → New: 0 thru 24 → 1 25 thru 49 → 2 50 thru 74 → 3 75 thru 100 → 4
O Rang <u>e</u> , value through HIGHEST:	☐ Output varia <u>b</u> les are strings/idth: 8
O All other values	Convert numeric strings to numbers ('5'->5)

Figure 4.8: Define New Variables

4.3 Steps for Conducting Descriptive Statistics in SPSS

a) Frequencies

We click the following: Analyze -> Descriptive Statistics -> Frequencies.

<u>File</u> Edit	View Data	Transform	Analyze	Graphs	Utilities	Extension	s <u>W</u> indow	Help			
F	A 🗖	10-3	Powe	Analysis		>		A 🕞	. 0		
- 10		-	Meta	Analysis		>			• •	•	
			Repor	ts		>					
	n 🖉 🕼	💑 Q2	Descr	iptive Stat	istics	>	Erequenci	es	~	뤚 Q14	SCODI_C
1	69	2	Bayes	sian Statis	tics	>	Descriptiv	es		1	4
2	45	2	Table	в		>	Decudation	Descriptions		1	5
3	59	1	Comp	are Means	5	>	Population	Descriptives		1	5
4	70	2	Gana	al Linear I	Model	,	A Explore			1	-3
5	43	2	Quine	al cilical i	nouel		Crosstabs	L		1	3
6	83	2	Gener	alized Line	ear models		TURF Ana	alysis		2	3
7	45	2	Mixed	Models		>	Patio			-1	-3
8	54	1	Correl	ate		>	Mano			2	-3
9	46	2	Regre	ssion		>	Proportion	Confidence II	ntervais	1	5
10	-1	1	Loglin	ear		>	P-P Plots			2	3
11	54	1	Neura	Networks	5	>	SQ-Q Plots			1	5
12	46	1	Class	ifv		>	1	2	1	1	3
13	67	1	Dime	uj nien Dedi	intian		1	2	2	1	3
14	-1	2	Quine	iaion reduc	JCtion .		1	3	3	1	5
15	81	2	Scale			,	4	2	1	1	5
16	51	2	Nonpa	irametric 1	lests	>	1	3	2	1	1
17	65	2	Forec	asting		>	1	3	3	1	3
18	64	2	Surviv	al		>	4	2	1	1	5
19	-1	-1	Multip	le Respon	ise	>	4	2	1	1	4
20	59	2	Missir	o Value A	nalysis		4	4	3	2	! 1
21	53	1	Adultic	la langutat	ion		-1	4	1	1	5
22	49	1	Comp	lex Same	lon	Ś	2	2	1	1	3
Data View	Variable View		Simula Out	ation		ĺ					

Figure 4.9: First Step in Choosing Frequencies Function

If you also want to display the variables as an illustration, click *Charts*, select the type of chart (Bar, Pie, Histogram) and specify whether you want to show frequencies or percentages.

ta Frequencies	×	Frequencies: Charts ×
[▲] Monitor your bloo [▲] Understand if you [▲] Understand if you [▲] Confidence [▲] Persist in monitor [▲] Take action to adj [▲] Evaluate if your a [▲] Persist in carryin [▲] Confidences [▲]	es_descri Statistics Charts Format Style Bootstrap	Chart Type O None Bar charts O Pie charts O Histograms: Show normal curve on histogram Chart Values
Display frequency tables	style tables	● Erequencies ○ Percentages
OK Paste Reset Canc	el Help	Continue Cancel Help

Figure 4.10: Choosing Variables When Conducting Frequencies Function

Results are presented in two tables:

Frequencies

Statistics

Confidence_descriptve

N	Valid	101
	Missing	40

Confidence_descriptve

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low self-confidences	1	.7	1.0	1.0
	Insufficient self- confidences	3	2.1	3.0	4.0
	Adequate self- confidences	34	24.1	33.7	37.6
	Good self-cnfidence	63	44.7	62.4	100.0
	Total	101	71.6	100.0	
Missing	System	40	28.4		
Total		141	100.0		

Figure 4.11: Results When Conducting Frequencies Function

The results are shown using frequencies and percentages. The valid percentage and cumulative percentage are also shown. The valid percentage is the percentage of cases that have non-missing values for each category. The cumulative percentage is the calculation of the valid percentages for each category or for previous categories.

The results can be reported as follows: One participant (1.0%) had low selfconfidence, three participants (3.0%) had insufficient self-confidence, 34 participants (33.7%) had adequate self-confidence, and 63 participants (62.4%) had high self-confidence.

The results can also be plotted using graphs.



Figure 4.12: Results in Graphs When Conducting Frequencies Function

b) Descriptives

We click on *Analyze* -> *Descriptive Statistics* -> *Descriptives*.

<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata	Transform	Analyze	Graphs	<u>U</u> tilities	Extensions	Window	<u>H</u> elp		
					Po <u>w</u> e	r Analysis		>		_ A 🖉		
_		9			Meta	Analysis		>	• ±±±± (<u> </u>	•
					Repor	ts		>				
		450 I	Q1	💑 Q2	Desci	iptive Stati	stics	>	Erequen	cies of	~	🚓 Q14
1			69	2	Baye	sian Statist	ics	>	Descript	ives		1
2	2		45	2	Table	s		>	Develop	Descriptions		1
3			59	1	Comr	are Means		,	Populati	on Descriptives		1
4			70	2	Cono	rol Lincor M	Indal		A Explore.			1
5			43	2	Gene		louel	í.	🐺 <u>C</u> rosstat	bs		1
6	;		83	2	Gene	ralized Line	ar Models	,	+ TURF Ar	nalysis		2
7			45	2	Mi <u>x</u> eo	I Models		>	Patio			-1
8			54	1	Corre	late		>	<u>IX</u> auo			2
9)		46	2	Regre	ssion		>	+ Proportio	on Confidence In	tervals	1
1)		-1	1	Loglin	ear		>	2-P Plot	8		2
1	1		54	1	Neura	I Networks		>	🛃 Q-Q Plot	ts		1
1	2		46	1	Class	ify		>	1	2	1	1
1	3		67	1	Dimo		ction		1	2	2	1
1	4		-1	2		ISION Redu	ction		1	3	3	1
1	5		81	2	Sc <u>a</u> le			,	4	2	1	1
1	5		51	2	Nonp	arametric T	ests	>	1	3	2	1
1	7		65	2	Forec	asting		>	1	3	3	1
1	8		64	2	Surviv	al		>	4	2	1	1
1	Э		-1	-1	Multip	le Respon	se	>	4	2	1	1
2	0		59	2	🜌 Missir	na Value Ar	nalvsis		4	4	3	2
2	1		53	1	Multir	le Imputati	00	>	-1	4	1	1
2	2		49	1	Comr	lov Samel		,	2	2	1	1
Data	View	Varial	ale View		Bimul	ation						
Data	*iew	- arron			Qualit	v Control		>				

Figure 4.13: First Step in Choosing Descriptives Function

Then click on Statistics, and we can choose different statistical parameters, such as Mean, Minimum, Maximum, Standard deviation, etc.

		Descriptives: Options ×
		<u>Mean</u> <u>S</u> um
Secriptives Variable(s):	× Options	Dispersion
Genter, [Q2] Genter, [Q3] Are you employed? [Q4]	Style <u>B</u> ootstrap	Distribution
Image: System constraints Image: System constraints Image: System constraints Image: System constraints		Display Order Varia <u>b</u> le list <u>A</u> lphabetic As <u>c</u> ending means Descending means
Save standardized values as variables		
OK Paste Reset Cancel Help		<u>Continue</u> Cancel Help

Figure 4.14: Choosing Variables When Conducting Descriptives Function

The results can be reported as follows: The average age of participants is 63.08 years (SD=12.96), with a minimum age of 24 years and a maximum age of 87 years (**Fig. 4.15**).

Descriptives

Descriptive Statistics								
	N	Minimum	Maximum	Mean	Std. Deviation			
Age (years):	132	24	87	63.08	12.961			
Valid N (listwise)	132							

Figure 4.15: Results When Conducting Descriptives Function

c) Explore

When choosing the function *Explore*, we can also choose different plots (e.g., Histogram, Stem and leaf, etc.) to present our data.

i		53	Power Analysis	>	-								
			Meta Analysis	>								Visible:	32 of 32 Variable
	A 01	A 02	Reports	>			~ · ·	A 014	SCODI C		2000	SCODI C	SCODI C S
	· 41	0 42	Descriptive Statistics	>	Erequencies.			00 02 14		2	3	4	5
1	69	2	Bayesian Statistics	>	Descriptives.			1	1 4	4	4	4	4
2	45	2	Tables	>	E Population D	accrintivas		1	1 5	5	5	5	3
3	59	1	Compare Means	>	T operation D	a a criptire a		1	1 5	3	2	5	5
4	70	2	General Linear Model	>	A Explore			1	1 -3	-3	-3	-3	-3
5	43	2	Generalized Linear Models	,	Crosstabs			1	1 3	3	1	3	5
6	83	2	Generalized Entear Models	÷	TURF Analys	is		2	2 3	3	4	5	4
7	45	2	Mixed Models	,	Ratio			-1	1 -3	-3	-3	-3	-3
8	54	1	Correlate	>				2	2 -3	-3	-3	-3	-3
9	46	2	Regression	>	Proportion Co	infidence int	ervais	1	1 5	5	5	3	5
10	-1	1	Loglinear	>	P-P Plots			2	2 3	3	3	5	5
11	54	1	Neural Networks	>	G-Q Plots			1	1 5	5	3	5	4
12	46	1	Classify	>	1	2	1	1	3	3	1	5	5
13	67	1	Dimension Reduction		1	2	2	1	1 3	1	1	4	4
14	-1	2	Dimension Reduction		1	3	3	1	1 5	5	3	5	5
15	81	2	Scale		4	2	1	1	1 5	4	4	5	4
16	51	2	Nonparametric Tests	>	1	3	2	1	1 1	1	3	5	5
17	65	2	Forecasting	>	1	3	3	1	1 3	2	3	2	4
18	64	2	Survival	>	4	2	1	1	1 5	5	5	5	5
19	-1	-1	Multiple Response	>	4	2	1	1	1 4	4	1	3	4
20	59	2	Missing Value Analysis		4	4	3	2	2 1	1	1	1	1
21	53	1	Multiple Imputation	>	-1	4	1	1	1 5	5	1	5	5
22	< 49	1	Complex Samples	>	2	2	1	1	1 3	3	2	4	5

Figure 4.16: First Step in Choosing Explore Function

4 Review of Statistical Basics

ta Explore	×	Explore: Plots	×
Confidences descript.	Statistics Plots Options Bootstrap	Boxplots Eactor levels together Dependents together None	Descriptive ☑ <u>S</u> tem-and-leaf ☑ <u>H</u> istogram
		Normality plots with tests Spread vs Level with Levene None O Power estimation	Test
Display ● <u>B</u> oth ○ St <u>a</u> tistics ○ P <u>l</u> ots		O Iransformed Power: Nat	ural log
OK Paste Reset Cancel Help		<u>Continue</u> Cance	Help

Figure 4.17: Choosing Variables When Conducting Explore Function

Results are presented in three tables:

Case Processing Summary

				Cas	ses		
		Va	lid	Miss	sing	Total	
	Confidence_descriptve	N	Percent	N	Percent	N	Percent
Age (years):	Low self-confidences	1	100.0%	0	0.0%	1	100.0%
	Insufficient self- confidences	3	100.0%	0	0.0%	3	100.0%
	Adequate self- confidences	31	91.2%	3	8.8%	34	100.0%
	Good self-cnfidence	60	95.2%	3	4.8%	63	100.0%

Figure 4.18: Summary Result When Conducting Explore Function

Skewness is a measure of symmetry. Kurtosis is a measure of the distribution. When the mean, median and mode coincide, it is called a symmetric distribution, i.e. skewness = 0, kurtosis = 0. Skewness and Kurtosis parameters must lie between 1 and -1 for the distribution to be approximately normal. This is a less reliable method for small to medium-sized samples (i.e. n < 300) because it cannot adjust the standard error. In this case, we apply a z-test using skewness and kurtosis (Mishra, et al., 2019).

Age (years): Insufficient self- confidences Mean 73.67 2.728 95% Confidence Interval for Mean Lower Bound 61.93 95% Confidence Interval for Mean Upper Bound 85.41 5% Trimmed Mean 72.00 Variance 22.333 Std. Deviation 4.726 Maximum 70 Range 9 Range 9 Kurtosis . 1.300 1.225 Kurtosis . . 5% Confidence Interval for Mean Lower Bound 66.35 2.075 5% Trimmed Mean 64.64 6% Confidence Interval for Mean Upper Bound 68.59 5% Trimmed Mean 64.64 Median 67.00 Variance 133.437 <th></th> <th>Confidence_descriptve</th> <th></th> <th></th> <th>Statistic</th> <th>Std. Error</th>		Confidence_descriptve			Statistic	Std. Error	
Sourtidences 95% Confidence Interval for Mean Lower Bound 61.93 1000 00000 85.41 100000 5% Trimmed Mean 72.00 100000 Variance 22.333 100000 Variance 22.333 100000 Variance 22.333 1000000 Variance 22.333 1000000000 Maximum 7000000000000000000000000000000000000	Age (years):	Insufficient self-	Mean		73.67	2.728	
for MeanUpper Bound85.415% Trimmed Mean.Median72.00Variance22.333Std. Deviation4.726Minimum70Maximum79Range9Interquartile Range.Skewness1.3901.225KurtosisKurtosis.95% Confidences64.3595% Confidence Interval for MeanLower Bound664.352.07595% Confidence Interval for Mean0.125% Trimmed Mean64.64Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness.623421Kurtosis.6234221.14Skewness.623421Kurtosis.503		confidences	95% Confidence Interval	Lower Bound	61.93		
5% Trimmed Mean.Median72.00Variance22.333Std. Deviation4.726Minimum70Maximum79Range9Interquartile Range.Skewness1.3901.225KurtosisSolo Confidence IntervalLower Boundfor Mean64.4352.07595% Confidence Intervalfor Mean64.64Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness623AdageSkewnessKurtosisSkewnessSkewnessSkewnessKurtosisStalStewnessSkewnessStalSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewnessSkewness			for Mean	Upper Bound	85.41		
Median72.00Variance22.333Std. Devlation4.726Minimum70Maximum79Range9Interquartile Range1.390Skewness1.3901.225Kurtosis1.390Skewness1.3901.225Skewness1.390Skewness1.3901.225Skewness1.390Skewness64.352.07595% Confidence Interval 107 MeanLower Bound664.64Median67.00Variance133.437Std. Devlation11.551Minimum41Maximum83Range42Interquartile Range14Skewness-623.421Kurtosis-503.821			5% Trimmed Mean				
Variance22.333Std. Deviation4.726Minimum70Maximum79Range9Interquartile Range1.390Skewness1.390Kurtosis.Kurtosis.95% Confidence Interval for MeanLower Bound664.352.07595% Confidence Interval for MeanLower Bound664.64.Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Maximum83Range42Interquartile Range14Skewness623421			Median		72.00		
Std. Deviation4.726Minimum70Range9Interquartile Range1.390Skewness1.390Kurtosis.Kurtosis.S% Confidence Interval for MeanLower Bound664.352.07595% Confidence Interval for MeanLower Bound664.64.Median66.64Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Maximum83Range14Skewness623Skewness623421KurtosisKurtosis503821			Variance		22.333		
Minimum70Maximum79Range9Interquartile Range1.390Skewness1.390Kurtosis.Kurtosis.Mean64.352.07595% Confidence Interval for MeanLower Bound60.125% Trimmed Mean64.64Median67.005% Trimmed Mean64.64Median67.00Variance133.4375td. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness.623.623.421Kurtosis.503.623.821			Std. Deviation		4.726		
Maximum79Range9Interquartile Range1.390Skewness1.390Kurtosis.Adequate self- confidencesMean95% Confidence Interval for MeanLower Bound64.352.07595% Confidence Interval for MeanLower Bound66.59.5% Trimmed Mean64.64Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness.623.623.421Kurtosis.503.503.821			Minimum		70		
Range9Interquartile Range1.3901.225Skewness1.3901.225Kurtosis			Maximum		79		
Interquartile Range1.3901.225Skewness1.3901.225Kurtosis			Range		9		
Skewness1.3901.225KurtosisAdequate self- confidencesMean64.352.07595% Confidence Interval for MeanLower Bound60.1210pper Bound68.595% Trimmed Mean64.64Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14KurtosisKurtosisKurtosisKurtosis			Interquartile Range				
KurtosisAdequate self- confidencesMean64.352.07595% Confidence Interval for MeanLower Bound60.1210Upper Bound68.5964.6464.6466Median67.0067.0011.55111.551Std. Deviation11.55111.55111.551Minimum418364.2466Range42141414Skewness623.421.421Kurtosis503.821.623.821	-		Skewness	Skewness			
Adequate self- confidencesMean64.352.07595% Confidence Interval for MeanLower Bound60.1210100 Per Bound68.5964.6464.64Median67.0067.0011.551Variance133.4375td. Deviation11.551Minimum414141Maximum838311Range421414Skewness623.421Kurtosis503.821			Kurtosis				
contridences95% Confidence Interval for MeanLower Bound60.12Upper Bound68.595% Trimmed Mean64.64Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness623.623.421Kurtosis503.821		Adequate self-	Mean	64.35	2.075		
for Mean Upper Bound 68.59 5% Trimmed Mean 64.64 Median 67.00 Variance 133.437 Std. Deviation 11.551 Minimum 41 Maximum 83 Range 42 Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821		confidences	95% Confidence Interval	Lower Bound	60.12		
5% Trimmed Mean 64.64 Median 67.00 Variance 133.437 Std. Deviation 11.551 Minimum 41 Maximum 83 Range 42 Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821			for Mean	Upper Bound	68.59		
Median67.00Variance133.437Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness623.421Kurtosis503.821			5% Trimmed Mean		64.64		
Variance 133.437 Std. Deviation 11.551 Minimum 41 Maximum 83 Range 42 Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821			Median		67.00		
Std. Deviation11.551Minimum41Maximum83Range42Interquartile Range14Skewness623Kurtosis503.821			Variance		133.437		
Minimum41Maximum83Range42Interquartile Range14Skewness623Kurtosis503.821			Std. Deviation		11.551		
Maximum 83 Range 42 Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821			Minimum		41		
Range 42 Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821			Maximum		83		
Interquartile Range 14 Skewness 623 .421 Kurtosis 503 .821	_		Range		42		
Skewness 623 .421 Kurtosis 503 .821			Interquartile Range		14		
Kurtosis503 .821			Skewness		623	.421	
			Kurtosis		503	.821	

Descriptives^a

Figure 4.19: Descriptives Results With the Explore Function

For small samples (less than 50 units), the Shapiro-Wilk test is recommended, although it can be used for larger samples, whereas the Kolmogorov-Smirnov test can only be used for larger samples (more than 50 units). If the p-value is higher than 0.05, we can confirm the normal distribution (Mishra, et al., 2019).

The results can be interpreted as follows: The Shapiro-Wilk test shows that the age distribution is normally distributed in all three groups (Fig. 4.20).

		Kolm	ogorov-Smir	nov ^b		Shapiro-Wilk	
	Confidence_descriptve	Statistic	df	Sig.	Statistic	df	Sig.
Age (years):	Insufficient self- confidences	.304	3		.907	3	.407
	Adequate self- confidences	.135	31	.158	.934	31	.056
	Good self-cnfidence	.075	60	.200	.983	60	.589

Tests of Normality^a

*. This is a lower bound of the true significance.

a. Age (years): is constant when Confidence_descriptve = Low self-confidences. It has been omitted.

b. Lilliefors Significance Correction

Figure 4.20: Tests of Normality When Conducting Explore Function

The data distribution can also be viewed by looking at the histogram. The histogram is an estimate of the probability distribution of the continuous variable. If we have a bell-shaped and symmetrical graph about the mean, we can assume that the data are normally distributed (Barton & Peat, 2014; Mishra, et al., 2019). In Figure 4.21, we see the abnormal age distribution in people with insufficient and sufficient self-confidence.







The Q-Q normal plot (**Fig. 4.22**) shows the relationship between two sets of quantiles (observed and expected). For normally distributed data, the observed data are approximately equal to the expected data.





Figure 4.22: Q-Q Normal Plot

We can also assess the normality of the data distribution using a box plot (**Fig. 4.23**). It shows the median as a horizontal line inside the box and the IQR (the range between the first and third quartiles) as the length of the box.



Figure 4.23: Box Plot

d) Crosstabs

Crosstabs can used to describe the relationship between two categorical variables. We click the following: *Analyze -> Descriptive Statistics -> Crosstabs*.

Eile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	Transform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	<u>W</u> indow	ı <u>H</u> elp		
					Po <u>w</u> er	Analysis		>				
					Meta	Analysis		>				
					Repor	ts		>				
		45 9 (Q1	💑 Q2	Descr	iptive Statis	stics	>	123 Ereque	ncies	A	🗞 Q14
1			69	2	Bayes	ian Statist	ics	>	Descri	otives		1
2	!		45	2	Ta <u>b</u> les	3		>	·	tion Descript	ivee	1
3			59	1	Comp	are Means		>		non Descript	IVES	1
4			70	2	Gener	al Linear M	lodel	>	A Explore	e		1
5			43	2	Gener	alized Line	ar Modele	\$	💥 <u>C</u> rosst	abs		1
6	;		83	2	Mined	Madala	al Wouers	,	+ TURF	Analysis		2
7			45	2	IVII <u>x</u> ed	woders			Ratio			-1
8			54	1	<u>C</u> orrel	ate		>	- Propor	tion Confiden	co Intonvolo	2
9)		46	2	<u>R</u> egre	ssion		>		coniden	ce intervais	1
1	0		-1	1	L <u>og</u> lin	ear		>	<u>ы Б-</u> ь ы	ots		2
1	1		54	1	Neura	l Networks		>	🛃 <u>Q</u> -Q PI	ots		1
13	2		46	1	Classi	fy		>	1	2	1	1
13	3		67	1	Dimer	ision Redu	ction	>	1	2	2	1
14	4		-1	2	Scale			>	1	3	3	1
1	5		81	2	Nonos	rametric T	acto	>	4	2	1	1
1	6		51	2	<u>In</u> onpa		6313		1	3	2	1
1	7		65	2	Forec	asting		,	1	3	3	1
1	B		64	2	Surviv	al		>	4	2	1	1
1	9		-1	-1	M <u>u</u> ltip	le Respons	se	>	4	2	1	1
2	0		59	2	🔣 Missin	g <u>V</u> alue Ar	nalysis		4	4	3	2
2	1		53	1	Mult <u>i</u> p	le Imputatio	on	>	-1	4	1	1
2:	2	<	49	1	Comp	ex Sample	s	>	2	2	1	1
Data	View	Variat	ole View		Bimula	ation						
Data	1.8W				Qualit	v Control		>				

Figure 4.24: Crosstabs

🔚 Crosstabs		×	n Crosstabs: Cell Display	>
 ✓ Age (years): [Q1] ✓ Education: [Q3] ➢ Are you employed? [Q4] ➢ What is your marital st ➢ Des anyone in your fa 	Rgw(s): Gender: [Q2] Column(s): Confidences_descriptive	Exact Statistics Cells Eormat	Counts Observed Expected Hide small counts Less than 5	Z-test
How would you rate yo How would you rate yo Do you think you have Check your blood suga When you have abnor When you have abnor When you have symmt	Layer 1 of 1 Previous Next	Style Bootstr <u>a</u> p	Percentages □ <u>Row</u> □ <u>C</u> olumn ☑ <u>I</u> otal □ Create A <u>P</u> A style table	Residuals Unstandardized Standardized Adjusted standardized
Find you have symptomic symptomi	Display layer variables in table laye	ers	Noninteger Weights	O Round case <u>w</u> eights O Truncate case weig <u>h</u> ts
OK <u>F</u>	Paste Reset Cancel Help		Continu	e Cancel Help

Figure 4.25: Choosing Variables When Conducting Crosstabs

The results can be seen in Figure 4.26:

In this table, the columns represent the values of the first variable, and the rows represent the values of the second variable. Cross-tabulations allow us to show the proportion of cases in subgroups.

				Confidence_descriptve								
			Low self- confidences	Insufficient self- confidences	Adequate self- confidences	Good self- cnfidence	Total					
Gender:	Men	Count	0	2	17	25	44					
		% of Total	0.0%	2.0%	17.0%	25.0%	44.0%					
	Women	Count	1	1	17	37	56					
		% of Total	1.0%	1.0%	17.0%	37.0%	56.0%					
Total		Count	1	3	34	62	100					
		% of Total	1.0%	3.0%	34.0%	62.0%	100.0%					

Gender: * Confidence_descriptve Crosstabulation

Figure 4.26: Results When Conducting Crosstabs

e) Compare Means

Compare means allows a comparison between numerical variables. We click the following: *Analyze -> Compare Means -> Means*.

File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extension	ns <u>W</u> in	dow Hel	p			
			_		Powe	r Analysis	-	- >		- -)	
_	1000		_		Meta	Analysis		>	-	। 14	C	. 💽 🧹		
					Reno	rte		>						
		🛷 Q*	1	💑 Q2	Desc	riptive Stati:	stics	>	<mark>6</mark> Q9	💑 Q1	10	💑 Q11	💰 Q1	14
6	1		40	2	Baye	sian Statist	ics	>		1	1	1		
6	2		28	2	Table	s		>		2	1	1		
6	3		48	2	Com	oare Means		>						12
6	4		34	2	Gana	ral Linear M	lodol		Me	ans				
6	5		40	2	Gene	rai Linear iv	iodei		C On	e- <u>S</u> ample T	Test.			
6	6		79	2	Gene	ralized Line	ar wodels	,	🛃 Ind	ependen <u>t</u> -S	ample	es T Test		
6	7		-1	2	Mixed	d Models		>	+ Su	mmary Inde	pende	ent-Samples	T Test	
6	B		42	1	Corre	late		>	Pai	red-Sample	e T T	oct		
6	Э		57	1	Regre	ession		>						
7	0		70	1	Loglin	near		>	<u>o</u> n	e-Way ANG	JVA			
7	1		72	1	Neura	al Networks		>	🛛 On	e-Sample F	roport	tions		
73	2		71	2	Class	ifv		>	🚮 Ind	ependent-S	ample	s Proportions	s	
7	3		62	2	Dime	nsion Redu	ction	,	🔛 Pai	red-Sample	es Pro	portions		
74	4		58	1	Cash		cuon		-	3	- 1			
7	5		56	2	Scale				-	1	2	3		
7	6		67	2	Nonp	arametric T	ests	>		1	3	1		
7	7		72	1	Forec	asting		>		1	3	3		
7	В		71	2	Surviv	/al		>		1	2	1		
7	9		79	1	Multip	ple Respons	se	>		1	2	1		
8	D		62	1	🕵 Missi	ng <u>V</u> alue Ar	nalysis		-	1	2	3		
8	1		87	2	Multi	ole Imputati	00	>	-	1	2	1		
8:	2		29	1	Com	lov Somele			-	1	-1	-1		-
-					Comp	Jex Cample	53				•••			
Data	View	Variable	View		Simul	ation								
-	_				Quali	ty Control		>						

Figure 4.27: Compare Means

			ta Means: Options	×
Means Follow advice a A Take your medi Persist in follow Monitor your blo Monitor your blo Cuderstand if y Recognize the Persist in monit Take action to a Persist in carryi Confidences_de	Dependent List: Age (years): [Q1] ayer 1 of 1 Previous Next Independent List: Confidences	× Qptions Style Bootstrap	Statistics: Median Grouped Median Std. Error of Mean Sum Minimum Maximum Range First Last Variance Kurtosis Std. Error of Kurtosis Std. Error of Kurtosis Std. Error of Skewnes: ↓ < Std. Error of Skewnes: ↓ Std. Error of Skewnes: ↓ Std. Error of Skewnes: ↓ Std. Error of Skewnes: ↓ Std. Error of Skewnes: ↓	Cell Statistics: Mean Number of Cases Standard Deviation
OK	Paste Reset Cancel Help		<u>Continue</u> Cance	l Help

Figure 4.28: Choosing Variables When Conducting Comparison Means

The results are shown in the following table:

Report									
Variables Age (years):									
confidencesscore	Mean	N	Std. Deviation						
22.73	59.00	1							
31.82	72.00	1							
36.36	79.00	1							
47.73	70.00	1							
50.00	67.00	2	7.071						
52.27	60.50	2	27.577						
54.55	59.00	3	13.528						
56.82	76.00	1							
59.09	45.00	1							
61.36	66.25	4	5.737						
63.64	71.50	2	.707						
65.91	68.00	2	2.828						
68.18	66.50	4	11.328						
70.45	66.20	5	14.025						
72.73	60.20	5	12.677						
75.00	65.57	7	11.844						
77.27	71.50	8	9.366						
79.55	70.29	7	10.515						
81.82	74.00	1							
84.09	54.50	4	9.983						
86.36	64.57	7	13.227						
88.64	70.00	3	15.133						
90.91	66.00	5	9.849						
93.18	60.67	6	11.361						
95.45	68.25	4	4.500						

Figure 4.29: Result when Conducting Compare Means

Custom Tables

f) Custom Tables

This is a frequency table that allows comparisons between variables. We click the following: *Analyze -> Tables -> Custome Tables*.

_			and the second se	Comite Trabue Tunnee	-Arension:	s <u>vv</u> indo	w Helb		
		A .		Power Analysis	>				2
-	1000			Meta Analysis	>		14 0		
				Reports	>				
		🦑 Q1	💑 Q2	Descriptive Statistics	>	6 Q9	💑 Q10	💑 Q11	💑 Q14
61		40	2	Bayesian Statistics	>	1	1	1	1
62		28	2	Ta <u>b</u> les	>	E Custo	m Tables	1	1
63		48	2	Compare Means	>	EE Multin	la Deenenee	Coto 1	-1
64		34	2	General Linear Model	>	To Multip	ie Re <u>s</u> ponse -	1	2
65		40	2	Generalized Linear Medele		E Define	Category Ord	der 1	1
66		79	2	Generalized Linear Woders		-1	2	1	1
67		-1	2	Mixed Models	,	1	2	1	2
68		42	1	Correlate	>	1	5	2	2 2
69		57	1	Regression	>	1	2	2	2 2
70		70	1	Loglinear	>	-1	1	1	1
71		72	1	Neural Networks	>	-1	1	1	1
72		71	2	Classify	>	-1	1	3	1
73		62	2	Dimension Reduction	,	1	2	1	1
74		58	1	Dimension Reduction		3	1	1	1
75		56	2	Scale	,	-1	2	3	1
76		67	2	Nonparametric Tests	>	1	3	1	1
77		72	1	Forecasting	>	-1	3	3	1
78		71	2	Survival	>	1	2	1	1
79		79	1	Multiple Response	>	1	2	1	1
80		62	1	Missing Value Analysis		-1	2	3	1
81		87	2	Multiple Imputation		-1	2	1	1
82		29	1	Complex Samples	,	-1	-1	-1	il -1
				Cimulation	,				
Data \	View	Variable Vie	9W	ug onnulauon					

Figure 4.30: Custom Tables

ables:						III INORM <u>a</u> l		ct	
Prevent high o		(Column	s			
Follow advice 🗸					Confidences	descriptive			
egories:				Low self	Insufficient	Sufficient	High self		
ow self-confidence				Count	Count	Count	Count		
nsufficient self-confi			Men	nnnn	nnnn	nnnn	nnnn		
Sufficient self-confide		O and an	Women	nnnn	nnnn	nnnn	nnnn		
ligh self-confidence		Gender:	Non-binary	nnnn	nnnn	nnnn	nnnn		
			Other:	nnnn	nnnn	nnnn	nnnn		
	Rows								
	Rows								
fine >	Rows	Summar	y Statistics						
fine % Summary Statistics	Rows	Summar Posjt	y Statistics ion: Colu	mns	~	∐ide	Category F	² osition:	

Figure 4.31: Choosing Variables When Conducting Custom Tables

The results are shown in Figure 4.32. In the figure, we can see the gender distribution of diabetes self-care and self-confidence levels.

Custom Tables

		Confidence_descriptve							
		Low self- confidences	Insufficient self- confidences	Adequate self- confidences	Good self- cnfidence				
		Count	Count	Count	Count				
Gender:	Men	0	2	17	25				
	Women	1	1	17	37				
	Non-binary gender	0	0	0	0				
	Other:	0	0	0	0				

Figure 4.32: Results When Conducting Custom Tables

g) Chart Builder

If you want to display the variables in a graph, click the following: *Graphs -> Chart Builder*.

ta Učb	enik_v1.	sav [Data	Set1] - IB	M SPSS Statis	tics Data Edito	r				
<u>F</u> ile	Edit	View	Data	Transform	<u>A</u> nalyze	Graphs	<u>U</u> tilities	Extensions	Window	He
					X	<mark>i C</mark> hart E	Builder			1
	_	-				Graph	board Tem	plate Chooser.		
		A c		0.00	0.02	🔀 Relatio	nship Ma	p	-	0.0
		« (21	øt Q2	6 0 Q3	+ Weibul	II Plot			e d.
	1		69	2	3.0	+ Compa	are Subgro	ups		
	2		45	2	3.0	+ Regres	sion Varia	able Plots		
	3		59	1	4.0					
	4		70	2	3.0	Legac	y Dialogs		>	
1	5		43	2	4.0		1	1	4	
	6		83	2	2.0		3	4	3	
	7		45	2	3.0		1	1	-1	
1	8		54	1	2.0		1	1	4	
1	9		46	2	2.0		3	2	1	
1	0		-1	1	3.0		1	1	1	
1	1		54	1	2.0		1	1	4	
1	2		46	1	3.0		1	3	1	
1	3		67	1	3.0		3	1	1	
1	4		-1	2	5.0		1	4	1	
1	5		81	2	3.0		3	1	4	
	-				2.2		-	-		

Figure 4.33: Chart Builder



Figure 4.34: Choosing a Variable for the Chart

Figure 4.35 shows graphically the gender ordering of the level of self-confidence of patients with diabetes mellitus.



Figure 4.35: Gender Distribution of Self-Confidence Levels in the Chart

4.4 Statistical Inference

Statistical inference is the process of concluding populations or scientific truths from data. There are many modes of performing inference, including statistical modelling, data-oriented strategies, and explicit use of designs and randomization in analyses.

Statistical tests can also be chosen according to the variables involved in the hypothesis and their distribution (**Fig. 4.36**). It should be noted that the following table represents a simplified visualization of statistical tests that can be used to test the hypothesis given the constraints of normal or non-normal distribution.

	Distri	bution		
Variables	Normal (Gaussian)	Not normal		
2 descriptive	Hi-square test (Cro	osstabs; Chi-square)		
1 descriptive and 1	Independent samples t-test	Mann-Whitney U Test		
numeric	(Compare Means; Independent	(Nonparametric Tests; Legacy		
	samples t-test)	Dialogs; 2 Independents		
		Samples)		
2 numerical	Correlation (Correlate - Biv	Correlation (Correlate - Biv.		
(correlation)	Pearson coefficient)	Spearman's coefficient)		
2 numeric (average)	t-test dependent samples	Wilcoxonov test (Nonparametric		
	(Compare Means; Paired samples	Tests; Legacy Dialogs; 2 Related		
	t-test)	Samples)		

4.5 Performing Independent Samples t-Test in SPSS

If we have one numeric variable and one descriptive variable with two classes and the distribution is normal, we can use the Independent samples t-test. It is used to check whether the mean difference between two groups is statistically significant (Mishra, et al., 2019). To perform the Student t-test we should click the following: *Analyze -> Compare Means -> Independent-Samples t-test* (Fig. 4.37).

<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	Transform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	<u>W</u> indow	<u>H</u> elp			
					Po <u>w</u> er	Analysis		>				2	
_					Meta	Analysis		>					
					Report	ts		>					
			Q1	💑 Q2	D <u>e</u> scr	iptive Statis	stics	>	Q9	💑 Q10	💑 Q11	💑 Q14	
6	1		40	2	Bayes	sian Statist	cs	>	1	1	1	1	
6	2		28	2	Ta <u>b</u> les	6		>	2	1	1	1	
6	3		48	2	Comp	are Means		>	M Means			-1	
64	4		34	2	Gener	al Linear M	odel	>				2	
6	5		40	2	Gener	alized Line	ar Models	>	Une-Sa	ample i Test		1	
6	6		79	2	Mixed	Generalized Linear Wodels			👬 Indepe	nden <u>t</u> -Sample	es T Test	1	
6	7		-1	2	MI <u>x</u> ed	Mixed Models			🛨 Summary Independent-Samples T Test				
6	3		42	1	Correl	<u>C</u> orrelate		>	🔚 Paired-Samples T Test				
6	9		57	1	Regre	ssion		>	One-Way ANOVA				
7)		70	1	L <u>og</u> lin	ear		>				1	
7	1		72	1	Neura	l Networks		>	One-Sample Proportions			1	
7	2		71	2	Classi	ify		>	Independent-Samples Proportions			s 1	
73	3		62	2	Dimer	sion Redu	ction	>	Paired-Samples Proportions			1	
74	4		58	1	Scale			>	э	1	1	1	
7	5		56	2	Nonos	rametric T	aete	,	-1	2	3	1	
7	6		67	2	<u>In</u> onpa		5313	, i	1	3	1	1	
7	7		72	1	Foreca	as <u>t</u> ing			-1	3	3	1	
7	3		71	2	Surviv	al		>	1	2	1	1	
7	9		79	1	Multip	le Respons	e	>	1	2	1	1	
8)		62	1	💕 Missin	ig <u>V</u> alue Ar	alysis		-1	2	3	1	
8	1		87	2	Multip	le Imputatio	on	>	-1	2	1	1	
8	2	<	29	1	Comp	lex Sample	s	>	-1	-1	-1	-1	
Data	View	Variat	ole View		🖶 Simula	ation				***			
_					Quality	Control		ς.					

Figure 4.37: First Step in Choosing Independent Samples t-Test

The numerical variable should be placed in the Test variable box, the ordinal variable should be placed in the Grouping variable box. We should also define groups, as presented:

Part And Anternation (1997) Test		×	
Age (years): [Q1]	Iest Variable(s):	Options Define Groups	×
Are you employed? [Q4]	<u>ب</u>	● <u>U</u> se specified values	
What is your mantal status? [Does anyone in your family h How would you rate your curr	-	Group <u>1</u> : 1	
How would you rate your com Do you think you have enough	<u>G</u> rouping Variable:	Group <u>2</u> : 2	
Check your blood sugar when	Q2(? ?)	O Cut point:	
When you have abnormal bloo When you have sumptons a OK Paste	Estimate effect sizes	<u>Continue</u> Cancel Help	

Figure 4.38: Choosing Variables When Conducting a t-Test

Group Statistics								
	Gender:	N	Mean	Std. Deviation	Std. Error Mean			
confidencesscore	Men	51	77.8075	16.93808	2.37180			
	Women	64	81.1080	16.47805	2.05976			

Results are presented in three tables:

	Independent Samples Test										
	Levene's Test for Equality of Variances				t-test for Equality of Means						
						Significance Mean		Mean	Std. Error	95% Confidence Differ	e interval of the rence
		F	Sig.	t	df	One-Sided p	Two-Sided p	Difference	Difference	Lower	Upper
confidencesscore	Equal variances assumed	.251	.618	-1.054	113	.147	.294	-3.30047	3.13150	-9.50453	2.9036
	Equal variances not assumed			-1.051	106.005	.148	.296	-3.30047	3.14135	-9.52849	2.9275

Independent Samples Effect Sizes										
			Point	95% Confide	ence Interval					
		Standardizer ^a	Estimate	Lower	Upper					
confidencesscore	Cohen's d	16.68317	198	566	.171					
	Hedges' correction	16.79493	197	562	.170					
	Glass's delta	16.47805	200	569	.170					

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.



First, we check for homogeneity of variance. To test for homogeneity of variances, we will use the results of Levene's test. If the p-value of Levene's test is greater than 0.05, it means that there are no differences in variances between the two groups and we print the results from the top row. A p-value for the Levene's test of less than 0.05 represents a significant difference between variances. We take the result from the row below.

Here is an example of how the results are reported: Based on Levene's test for homogeneity of variances, we found that the assumption of homogeneity of variances for self-confidence score was appropriate for male and female participants (p=0.618). The difference in mean self-confidence score between males and females is not statistically significant (M=-3.300; 95% CI [-9.505; 2.904]; p = 0.294).

4.6 Performing the Kruskal Wallis Test in SPSS

If you have one numeric variable and one descriptive variable with more than two classes, and the distribution is not normal, choose the Kruskal Wallis test. To perform the Kruskal Wallis test we should click the following: *Analyze -> Nonparametric Test -> Legacy Dialogs -> K Independent Samples*.

<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>D</u> ata	Transform	<u>Analyze</u> <u>Graphs</u> <u>Utilitie</u>	s E <u>x</u> tensions	Window	<u>H</u> elp					
			Power Analysis	>							
			Meta Analysis	>				•			
			Reports	>							
	🛷 Q1	💑 Q2	Descriptive Statistics	>	<mark>6</mark> Q9	💑 Q10	🗞 Q11	💑 Q14	SCODI_C		
61	40	2	Bayesian Statistics	>	1	1	1	1	5	5	3
62	28	2	Ta <u>b</u> les	>	2	1	1	1	-3	-3	-3
63	48	2	Compare Means	>	-1	2	1	-1	-1	-1	-1
64	34	2	General Linear Model	>	4	2	1	2	1	2	1
65	40	2	Ganaralizad Linear Mada		2	2	1	1	5	3	2
66	79	2	Generalized Linear Wode	15	-1	2	1	1	1	2	3
67	-1	2	MIXed Models	>	1	2	1	2	5	4	5
68	42	1	Correlate	>	1	5	2	2	-3	-3	-3
69	57	1	Regression	>	1	2	2	2	2	1	1
70	70	1	L <u>og</u> linear	>	-1	1	1	1	3	3	3
71	72	1	Neural Networks	>	-1	1	1	1	3	3	1
72	71	2	Classify	>	-1	1	3	<u>X</u> <u>C</u> h	ii-square		1
73	62	2	Dimension Reduction	>	1	2	1	0/1 <u>B</u> ir	nomial		1
74	58	1	Seele		3	1	1	Bu Ru	ins		3
75	56	2	OCTION TO THE REAL		-1	2	3				1
76	67	2	Nonparametric Tests	>	🛕 <u>O</u> ne Sa	mple		A 1-3	Sample K-S		5
77	72	1	Forecasting	>	// Indeper	ident Sample	S	2	ndependent Sa	mples	2
78	71	2	<u>S</u> urvival	>	Related	Samples		🚺 <u>K</u> I	Independent Sa	mples	1
79	79	1	Multiple Response	>		Nonnarametri		📉 2 F	Related Sample	s	3
80	62	1	Missing Value Analysis		- Quade	nonparametri	IC ANOUNA		Polated Sample		2
81	87	2	Multiple Imputation	>	Legacy	Dialogs			Relateu gampie		5
82	< 29	1	Complex Samples	>	-1	-1	-1	-1	-1	-1	-1
Data View	Variable View		Bimulation			•••					
			Quality Control	>							

Figure 4.40: First Step in Choosing Kruskal Wallis Test

The numerical variable should be placed in the Test variable box, the ordinal variable should be placed in the Grouping variable box. We should also define groups, based on how many classes are included:

talease for Several Independent Samples	×
 Are you employed? [Q4] What is your marital status? [Q5] Does anyone in your family hav How would you rate your comm Do you think you have enough When you have abnormal blod When you have abnormal blod When you have symptoms, an 	Exact Options Image for Grouping Variable Minimum:
Test Type	Ma <u>x</u> imum: 5
OK Paste Reset Cancel Help	<u>Continue</u> Cancel Help

Figure 4.41: Choosing Variables When Conducting Krusal Wallis Test

Results are presented in two tables:

	Ranks		
	How would you rate your current health?	N	Mean Rank
Confidences	Very good	12	78.75
	Good	67	64.10
	Sufficient	30	36.92
	Bad	4	46.75
	Very bad	1	21.00
	Total	114	

Kruskal-Wallis Test

Test Statistics^{a,b}

	Confidences					
Kruskal-Wallis H	20.999					
df	4					
Asymp. Sig.	<.001					
a. Kruskal Wallis Test						
b. Grouping Var would you rat current health	iable: How e your I?					

Figure 4.42: Results When Conducting Krusal Wallis Test

The results can be reported as follows: Based on the Kruskal Wallis Test, we found that an assessment of the participant's current state of health statistically affects the self-confidence score ($\chi 2 = 20.999$; p < 0.001).

4.7 Performing the ANOVA Test in SPSS

If you have one numeric variable and one descriptive variable with more than two classes, and the distribution is normal, choose the ANOVA test. To perform the ANOVA test, we should click the following: *Analyze -> Compare Means -> One-Way ANOVA*. We can also run a post-hoc test to see which variables are statistically significantly different.

<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	Transform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	<u>W</u> indow	<u>H</u> elp			
				P 3	Po <u>w</u> er	r Analysis		>					
	-	- Br			Meta	Analysis		>					
67 : S	CODI_C	21	5		Repor	ts		>					
		<u>چې</u> (21	💑 Q2	Descr	iptive Statis	stics	>	6 Q9	뤚 Q10	💑 Q11	💑 Q1	14
6	1		40	2	Bayes	sian Statist	ics	>	1	1	1		1
6	2		28	2	Ta <u>b</u> les	6		>	2	1	1		1
6	3		48	2	Comp	are Means		>	M Means				-1
6	4		34	2	Gener	al Linear N	lodel	>					2
6	5		40	2	Gener	alized Line	ar Models	,	Une- <u>s</u> a	mpie i rest.			1
6	6		79	2	Mined		ar modelo		🔛 Indepen	den <u>t</u> -Sample	es T Test		1
6	7		-1	2	MIXed	iviodels			🛨 Summa	ry Independe	ent-Samples 1	Test	2
6	8		42	1	Correl	ate		>	Baired-S	Samples T T	est		2
6	9		57	1	<u>R</u> egre	ssion		>					2
7	0		70	1	L <u>og</u> lin	ear		>		iy ANOVA			1
7	1		72	1	Neura	I Networks		>	Z One-Sa	mple Propor	tions		1
7	2		71	2	Classi	ifv		>	📓 Indepen	dent-Sample	es Proportions		1
7	3		62	2	Dimer	nsion Redu	ction	>	Paired-S	Samp <u>l</u> es Pro	portions		1
7	4		58	1	Scale			>	э	- 1	1		1
7	5		56	2	OC <u>a</u> le				-1	2	3		1
7	6		67	2	Nonba	arametric 1	ests	,	1	3	1		1
7	7		72	1	Forec	as <u>t</u> ing		>	-1	3	3		1
7	8		71	2	Surviv	al		>	1	2	1		1
7	9		79	1	M <u>u</u> ltip	le Respons	se	>	1	2	1		1
8	0		62	1	💋 Missin	ng <u>V</u> alue Ar	alysis		-1	2	3		1
8	1		87	2	Multin	le Imputati	on	>	-1	2	1		1
8	2	/	29	1	Comp	lov Samala		\$	-1	-1	-1		-1
		•			E o inp	Tev Cample				•••			
Data	a View	Variat	le View		Ten Simula	ation							
_					Qualit	v Control		>					

Figure 4.43: First Step in Choosing ANOVA Test

Then, we should choose one numerical and one descriptive variable.



Figure 4.44: Choosing Variables When Conducting ANOVA Test

Age (years):		ANOVA	ı		
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	677.959	2	338.980	2.124	.124
Within Groups	20109.544	126	159.600		
Total	20787.504	128			

Oneway

Figure 4.45: Results When Conducting ANOVA Test

The results can be reported as follows: Based on ANOVA, we found that the test is not statistically significant (F(2, 126)=338.980; p=0,124).

4.8 Performing the Correlation Test in SPSS

Correlations are used when we are measuring a linear relationship between two numerical variables. In the case of a normal distribution, we use the Pearson correlation coefficient, and in the case of an unequal distribution, we use the Spearman correlation coefficient.

To perform the correlation test we should click the following: *Analyze -> Correlate - > Bivariate*:

Eile	Edit	View	Data	Transform	Analyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	Window	<u>H</u> elp		
				P	Powe	r Analysis		>				2
				_	Meta	Analysis		>				N
67 : S	CODI_C	21	5		Rego	rts		>				
		450	Q1	💑 Q2	Desc	riptive Stat	istics	>	6 Q9	💑 Q10	💑 Q11	💑 Q14
7	3		62	2	Baye	sian Statis	tics	>	1	2	1	1
7	4		58	1	Ta <u>b</u> le	s		>	3	1	1	
7	5		56	2	Comp	are Means		>	-1	2	3	1
7	6		67	2	Gene	ral Linear M	Aodel	,	1	3	1	
7	7		72	1	Gono	ralized Lin	ar Modele	,	-1	3	3	1
7	8		71	2	Gene		ear mouers	, ,	1	2	1	1
7	9		79	1	wixed	1 Wodels		,	1	2	1	· · ·
8	0		62	1	Corre	late		>	🛨 Bivariat	e with Confid	ence Interval	S
8	1		87	2	<u>R</u> egre	ssion		>	🔢 Bivariat	e		
8	2		29	1	L <u>og</u> lir	lear		>	Partial.			
8	3		51	2	Neura	I Networks	3	>	Dictore	-		1.00
8	4		78	1	Class	ify		>	O Distant			
8	5		62	1	Dime	nsion Redu	uction	>	E Canonio	cal Correlatio	n	
8	6		58	2	Scale			,	1	1	3	1
8	7		63	1	Nege				2	2	3	1
8	8		83	1	<u>IN</u> ONP.	arametric	ests		-1	1	1	
8	9		74	2	Forec	asting		,	4	2	3	1
9	0		-1	2	Surviv	al		>	2	1	1	
9	1		-1	2	Multip	ole Respon	se	>	3	2	1	
9	2		63	2	🐉 Missir	ng <u>V</u> alue A	nalysis		1	2	1	
9	3		66	2	Multip	ble Imputat	ion	>	1	-1	1	1
9	4	<	75	1	Comp	lex Sampl	es	>	3	3	1	1
Data	n View	Varial	ble View		Bimul	ation						
_		_			0							

Figure 4.46: First Step in Choosing Correlation Test

Then, we should choose two numerical variables and correlation coefficient:

Bivariate Correlations	×
 ➢ Monitor your blood sugar as o ∧ ➢ Understand if your blood suga ➢ Recognize the symptoms of I ➢ Persist in monitoring your dia ➢ Evaluate if your actions were ➢ Persist in carrying out actions ➢ Confidences_descriptive 	Options Style Bootstrap Confidence interval
Correlation Coefficients ☑ PearsonKendall's tau-bSpearman	
Test of Significance ⊙ <u>T</u> wo-tailed O One-tailed	
Elag significant correlations Show only the lower triangle Show diagonal	
OK Paste Reset Cancel Help	

Figure 4.47: Choosing Variables When Conducting Correlation Test

Results are presented in the table (Fig. 4.48):

Correlations

Correlations

		Age (years):	Confidences
Age (years):	Pearson Correlation	1	076
	Sig. (2-tailed)		.430
	N	132	109
Confidences	Pearson Correlation	076	1
	Sig. (2-tailed)	.430	
	N	109	116

Figure 4.48: Results When Conducting Correlation Test

Figure 58 explains the correlation values:

Value	Explanation
-1	Perfect linear relationship (negative)
-0.70	Very strong correlation (negative)
-0.40	Medium or moderate correlation (negative)
-0.10	Very high or very strong downward correlation (negative)
0.00	No coherence
+0.10	Low or weak connectedness (positive)
+0.40	Medium or moderate (positive)
+0.70	Very strong correlation (positive)
+1	Full or functional connectivity (positive)

Figure 4.49: Explanation of Correlation Value

The results can be reported as follows: The Pearson's correlation test did not show a correlation between age and self-care confidence score. Age and self-care confidence scores are not statistically significantly correlated (rs=-0.076; p=0.430).

4.9 Performing the Chi-Square Test in SPSS

To compare two independent descriptive variables, we can use the Chi-square test and Fisher's exact test. The Chi-square test is applied assuming a large sample size, while the Fisher exact test is applied to small samples. The Fisher exact test is generally used for small samples, but can also be used for larger samples (Kim, 2017).

To perform the Chi-square test in the SPSS, we should click the following: *Analyze* -> *Descriptive Statistics* -> *Crosstabs*

<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>D</u> ata	<u>T</u> ransform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	E <u>x</u> tensions	<u>W</u> indow	<u>H</u> elp			
				r 3	Po <u>w</u> er	Analysis		>					
		- <u></u> -			Meta /	Analysis		>					
67 : 50		1	5		Report	ts		>					
		<u>م</u>	21	💑 Q2	D <u>e</u> scri	iptive Statis	stics	>	123 Erequei	ncies		💑 Q14	SCODI_C
9	7		60	2	Ba <u>v</u> es	ian Statist	ics	>	🛂 Descrip	tives		1	3
9	8		70	1	Ta <u>b</u> les	6		>	- Populat	ion Descripti	100	1	4
9	9		66	1	Comp	are Means		>		ion Descripti	ves	2	1
10	00		63	1	Gener	al Linear M	lodel	>	A Explore			1	3
10)1		59	2	Ganar	alizad Lina	or Modele		🐺 <u>C</u> rossta	ıbs		1	5
10)2		84	2	Minud		al Wouers	,	+ TURF A	nalysis		1	3
10)3		75	2	IVII <u>x</u> ed	woders		,	Ratio			1	3
10)4		80	1	<u>C</u> orrela	ate		>	- Droport	ion Confidenc	o Intonvolo	2	3
10)5		62	2	Regre	ssion		>	Fropon	ion Coniident	e intervais	1	4
10)6		74	2	Loglin	ear		>	2-P Plo 🔁	ts		1	3
10)7		76	1	Neural	l Networks		>	🛃 <u>Q</u> -Q Pli	ots		2	3
10	8		52	2	Classi	fy		>	-1	2	2	2	5
10)9		43	2	Dimen	ision Redu	ction	>	-1	2	1	1	5
11	0		72	2	Scale			>	-1	2	1	1	5
11	1		62	2	Nappa	romotrio T	ooto		-1	3	2	1	5
11	2		53	2	Nonpa	irametric 1	ests		-1	2	1	1	5
11	13		74	1	Foreca	asting		>	-1	3	1	1	5
11	4		71	1	<u>S</u> urviva	al		>	-1	2	1	1	5
11	15		68	2	M <u>u</u> ltip	le Respons	e .	>	-1	2	2	1	5
11	16		47	1	🏭 Missin	g <u>V</u> alue Ar	nalysis		-1	2	1	1	4
11	7		74	2	Multip	le Imputatio	on	>	-1	2	1	1	5
11	8	<	68	2	Comp	lex Sample	es.	,	1	2	1	1	5
Data	View	Variab	le View		Bimula Simula	ation							
					Quality	v Control		>					

Figure 4.50: First Step in Choosing Chi-Square Test

Both variables are ordinal, one should be placed in the Row and the other in the Column. Then we should click on Statistics to choose the Chi-square test:

Crosstabs	×	Crosstabs: Statistics	×
	Exact Statistics Calls Eormat Style Bootstrap	Chi-square Nominal Contingency coefficient Phi and Cramer's V Lambda Uncertainty coefficient Nominal by Interval Eta	Correlations Ordinal Gamma Somers' d Kendall's tau-b Kendall's tau-c Kappa Risk McNemar
Display layer variables in table layers Display clustered bar charts Suppress tables OK Paste Reset Cancel Help		Cochr <u>a</u> n's and Mantel-Hae	equals: 1 Help

Figure 4.51: Choosing Variables When Conducting Chi-Square Test

Results are presented in tables:

Gender: * Do you think you have enough knowledge about diabetes self-care and management? Crosstabulation

			Do you think you knowledge abou care and ma		
			Yes, I think I have sufficient knowledge	No, I don't think I have enough knowledge	Total
Gender:	Men	Count	44	14	58
		% of Total	32.6%	10.4%	43.0%
	Women	Count	64	13	77
		% of Total	47.4%	9.6%	57.0%
Total		Count	108	27	135
		% of Total	80.0%	20.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	1.088 ^a	1	.297		
Continuity Correction ^b	.682	1	.409		
Likelihood Ratio	1.079	1	.299		
Fisher's Exact Test				.385	.204
Linear-by-Linear Association	1.080	1	.299		
N of Valid Cases	135				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 11.60.

b. Computed only for a 2x2 table

Figure 4.52: Results When Conducting Chi-Square Test

The results can be reported as follows: Based on a Chi-square test, there is no static association between gender and diabetes knowledge (p=0.297).

4.10 Performing the Mann-Whitney U Test in SPSS

If you have one numeric variable and one descriptive variable with two classes and the distribution is not normal, use the Mann-Whiteney U test. To perform the Mann-Whiteney U test, we should click the following: *Analyze -> Nonparametric Tests -> Legacy Dialogs -> 2 Independent Samples*.



Figure 4.53: First Step in Choosing Mann-Whiteney U Test

The numerical variable should be placed in the Test variable box, the ordinal variable should be placed in the Grouping variable box. We should also define groups as presented:

ta Two-Independent-Samples Tests		×			
Education: [Q3] Carlor Are you employ What is your ma Does anyone in	Test Variable List: Confidences	Exact Options	tian Two Indep	endent Samples: De	×
Image: Second system Image: Second system Image: Second	<u>G</u> rouping Variable: Q14(1 2) Define Groups		Group <u>1</u> :	1	
Test Type ✓ <u>M</u> ann-Whitney U	Lolmogorov-Smirnov Z		Group 2:	2	
Moses extreme reactions	Wald-Wolfowitz runs		<u>C</u> ontinue	Cancel Help	

Figure 4.54: Choosing Variables When Conducting Mann-Whiteney U Test

Results are presented in tables:

NPar Tests

Mann-Whitney Test

Ranks				
	Do you think you have enough knowledge about diabetes self-care and management?	Ν	Mean Rank	Sum of Ranks
Confidences	Yes, I think I have sufficient knowledge	92	63.70	5860.50
	No, I don't think I have enough knowledge	22	31.57	694.50
	Total	114		

Test Statistics^a

	Confidences			
Mann-Whitney U	441.500			
Wilcoxon W	694.500			
Z	-4.105			
Asymp. Sig. (2-tailed)	<.001			
a. Grouping Variable: Do you think you have enough knowledge about diabetes self-care and management?				

Figure 4.55: Results When Conducting Mann-Whiteney U Test

The results can be reported as follows: The Mann-Whitney U test showed that there is a statistical effect of knowledge about diabetes self-care and management on confidences score (U=441.500; p<0.001).



5 Regression

This chapter introduces the characteristics of regression, with detailed discussions of linear regression, logistic regression and multiple regression. In addition to the theoretical background for each type of regression, practical guidance is provided on how to perform these regression analyses using SPSS software. The learning objectives of this chapter are to enable the reader to understand the basic concepts of the different types of regression, including their application and interpretation of results. The practical part of the chapter will teach the reader how to use SPSS to perform linear, logistic and multiple regression. This includes preparing the data, setting up the analyses and interpreting the output. In this way, the chapter will provide the reader with a comprehensive understanding of the different regression techniques and how they can be applied in practice.

Regression is a method by which we want to test the relationship between two numerical variables. We have univariate, bivariate, and multivariate regression analysis at our disposal. Regression is also distinguished according to the type of dependent variable: regression-numerical (normally distributed) and logistic regression (the dependent variable has a binomial distribution). Regression can also be linear or nonlinear (square, cubic, exponential, ...).

5.1 Linear Regression

Simple (simple, bivariate) linear regression tests the dependence of random variables. The regression function represents the regression line. The variable being predicted is called the dependent variable. The variable used to predict the other variable is called the independent variable. This relationship or model can be written in the form of an equation:

$$Y = \alpha + \beta \cdot X + \varepsilon$$

Where Y is the (dependent) variable or outcome studied, X is the (independent) explanatory variable or predictor, and ε is a random variable (error) whose role is:

- a) measurement (rounding) error;
- b) another random deviation from the linear relation $Y = \alpha + \beta X$.

We can assume $E(\varepsilon) = 0$. On average, then, a linear relation holds. However, we can also assume that $\varepsilon \sim N(0, \sigma)$ for some unknown deviation σ .

A sample of size n is n pairs of measurements (Xi, Yi). They are obtained by n replications of a random experiment.

The parameters of the model α and β are estimated so that the line Yb = α + β X fits the data as closely as possible.

The regression line (y') can be written as y' = a + bx. We can draw several regression lines between the points that will fit these points. It is necessary to determine the criteria that will determine which line best fits the given points.

The parameter a determines where the regression line intersects the ordinate, a = Y (0), so Y has the value of a when X has the value 0, b determines the slope of the line (positive or negative connection and connection strength). b is called the regression coefficient. It tells how much the value of Y changes if X changes by one unit. If b = 0, then Y does not depend on X (the variable Y is a constant for which the value of the variable X has the same value, i.e. Y '= a).
Most commonly, we use the least-squares method, which requires that the sum of the squares of the deviations from the regression line be minimal.

Performing regression analysis also includes certain steps: verification of certain assumptions, verification of the reasonableness of the model, and interpretation of regression coefficients.

Assumptions for performing linear regression are:

- Variables must be measured at a continuous level.
- The variables must be independently observable of each other.
- The residuals (errors) of the best-fitting regression line have a normal distribution.
- The ratio between X and Y is linear; a linear relationship between the two variables is expected). To check whether a linear relationship exists can create a scatterplot that may look like one of the following:



Figure 5.1: Linear Regression

 Y values are normally distributed around the regression line, and there are no significant outliers:



Figure 5.2: Outlier Regression

- The scattering of Y is constant around the regression line, and data show homoscedasticity:



Figure 5.3: Heteroscedasticity Regression

- observations are independent.

Before performing the analysis, the listed assumptions must be verified.

Graphic display

- Spread diagrams are used to show the relationship of two numerical variables.
 From the scattering graph, we can understand the pairs drawn as points in the coordinate system.
- Responsible surfaces.

We estimate the regression constant (a) and the regression coefficient (b) when interpreting the model.

To judge how good the regression model (regression function) is, we check:

- Determination coefficient (share of explained variance) a quality indicator of the description of the dependence between the variables with a regression line.
- Standard estimation error an indicator of the quality of predicting the value of a dependent variable using a regression line.

5.2 Steps to Conduct the Linear Regression in SPSS

To perform the linear regression, we should click the following: *Analyze ->* Regression -> *Linear*.

<u>F</u> ile	Edit	View	<u>D</u> ata	Transform	Analyze	Graphs	<u>U</u> tilities	Extensions	<u>W</u> indow	Help		
					Po <u>w</u> e	r Analysis		>			0	Q
					Meta	Analysis		>				
		A C	21	A 02	Repor	ts		>	2 09	A 010	A 011	A 014
		×	241	0 0 QZ	Desci	iptive Stati	stics	>	0 025	0 0 Q 10	an an	0 0 (214
	1		69	2	Bayes	sian Statist	ics	>	4	2	1	1
	2		45	2	Table	s		>	4	3	2	2
	3		59	1	Comp	are Means		>	-1	1	1	1
	4		70	2	Gene	ral Linear N	lodel	>	4	2	1	1
	5		43	2	Gene	ralized Line	ar Models	>	4	1	1	1
	6		83	2	Mixer	Models		,	3	3	1	3
	7		45	2	1411200				-1	2	1	-
	8		54	1	<u>C</u> orrel	late		,	4	3	1	4;
	9		46	2	Regre	ssion		>	Autom	atic Linear M	lodeling	
1	10		-1	1	Loglin	lear		>	🔣 Linear.			-
1	11		54	1	Neura	I Networks		>	Curve E	stimation		
1	12		46	1	Class	ify		>	Dertiel	Logat Cauge		-
1	13		67	1	Dime	nsion Redu	ction	>		Least Oquar	es	-
1	14		-1	2	Scale			>	Binary	Logistic		
1	15		81	2	Nonn	arametric T	ecte	,	🔛 Multino	mial Logistic	2	
1	16		51	2	<u>R</u> onpa	and the first in the second se	0000	,	Grdinal			
1	7		65	2	Forec	asing			Probit			
								-				

Figure 5.4: First Step When Choosing Linear Regression

Transfer the independent variable (e.g., confidences score), into the Independent(s) box and the dependent variable (e.g., age), into the Dependent box.

	Dependent:	Statistics
🔓 Gender: [Q2]	Age (years): [Q1]	Dista
B Education: [Q3]	Block 1 of 1	Plo <u>i</u> s
윩 Are you employed		S <u>a</u> ve
윩 What is your marit	Pre <u>v</u> ious <u>N</u> ext	Ontions
윩 Does anyone in yo	Block 1 of 1	<u>options</u> .
윩 How would you rat	Section Confidences	Style
윩 How would you rat		Bootstrap
윩 Do you think you h		
윩 Check your blood		
윩 When you have ab		
윩 When you have ab		
윩 When you have sy		
💫 If you find out that		
💫 If you find out that	Method: Enter Y	
윩 After taking action 🖳	Selection Variable:	
💫 If you find out that	Rule	
💑 Do you take insuli	,	
💫 If you find out that	Case Labels:	
Descent bisk as law.	★	
Prevent high or low		
Follow advice abou	WLS Weight:	

Figure 5.5: Choosing Variables When Conducting Linear Regression

To check the above stated assumptions, the researcher can use the Statistics and Plots features and then select the appropriate options within these two dialogue boxes.

Click on the OK button and generate the results.

The first table is the Model Summary table:

Model Summary										
Model	Std. Error of the Estimate									
1	.076 ^a	.006	003	11.256						
a Predictors: (Constant) Confidences										

Figure 5.6: Results (Model Summary) of Linear Regression

This table provides the R and R^2 values. The R value represents the simple correlation and is 0.076 (the "R" Column), indicating the variables are unrelated. If R is close to -1 it means that there is a negative correlation between the variables, if R is close to 0 it means that the variables are uncorrelated, if R is close to 1 it means that the variables are uncorrelated, if R is close to 1 it means that the variables are uncorrelated.

The R² value (the "R Square" column) indicates how much of the total variation is in the dependent variable. In this case, 0.6% can be explained as very low.



Figure 5.7: Results of Normal P-P Plot of Regression

fits the data:

ANOVA^a

The next table is the ANOVA table, which reports how well the regression equation

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	79.352	1	79.352	.626	.430 ^b
	Residual	13556.336	107	126.695		
	Total	13635.688	108			

a. Dependent Variable: Age (years):

b. Predictors: (Constant), Confidences

Figure 5.8: Results (ANOVA) of Linear Regression

This table indicates that the regression model predicts the dependent variable significantly well. How do we know this? Look at the "Regression" row and go to the "Sig." column. This value indicates the statistical significance of the regression model that was run. Here, p = 0.430, indicates that, overall, the regression model does not statistically significantly predict the outcome variable (i.e., it is not a good fit for the data).

The Coefficients table provides us with the necessary information to predict total confidence score from age, as well as determine whether income contributes statistically significantly to the model (by looking at the "Sig." column):

Coefficients ^a										
Unstandardized Coefficients Coefficients										
Model		В	Std. Error	Beta	t	Sig.				
1	(Constant)	69.277	5.295		13.084	<.001				
	confidencesscore	052	.065	076	791	.430				

a. Dependent Variable: Age (years):

Figure 5.9: Results (Coefficients) of Linear Regression

5.3 Logistic Regression

The logistic regression predicts the probability that a variable falls into one of two dichotomous dependent variable categories based on one or more independent variables. In the logistic regression, the following assumptions must be met:

- The dependent variable must be measured on a dichotomous scale.
- One or more independent variables, which may be continuous, ordinal or nominal, must be included.
- There must be an independent observation.
- There must be a linear relationship between all independent variables and a logit transformation of the dependent variables (Stoltzfus, 2011; Sperandei, 2014).

5.4 Steps to Conduct the Logistic Regression in SPSS

To perform the logistic regression, we use the following function: *Analyze ->* Regression -> Binary Logistic.

<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata	Transform	<u>A</u> nalyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	<u>W</u> indov	v <u>H</u> elp		
					Po <u>w</u> er	r Analysis		>		A (Q
_		5°			Meta	Analysis		>				~
					Repor	ts		>				
		🌮 Q	1	💑 Q2	D <u>e</u> scr	iptive Statis	stics	>	6 Q9	💑 Q10	💑 Q11	🗞 Q14
1	1		69	2	Bayes	sian Statist	ics	>	4	2		1 1
2	2		45	2	Ta <u>b</u> les	S		>	4	3	:	2 1
3	3		59	1	Comp	are Means		>	-1	1		1 1
4	1		70	2	Gener	ral Linear N	lodel	>	4	2		1 1
5	5		43	2	Gener	alized Line	ar Models	>	4	1		1 1
6	5		83	2	Mixed	Models		,	3	3	:	3 2
7	7		45	2	Comel	-t-			-1	2		1 -1
8	3		54	1	Correl	ate		,	4	3		1 2
9	9		46	2	Regre	ssion		>	Autom	atic Linear M	lodeling	1
1	0		-1	1	Loglin	ear		>	🚠 Linear.			2
1	1		54	1	Neura	I Networks		>	Curve	Estimation		1
13	2		46	1	Class	ify		>	B Partial	Least Squar	oc	1
1	3		67	1	Dimer	nsion Redu	ction	>	n arciai	Lea <u>s</u> r oqual	63	1
14	4		-1	2	Sc <u>a</u> le			>	Binary	Logistic		1
1	5		81	2	Nonpa	arametric T	ests	>	Rultine	omial Logistic	o	1
10	6		51	2	Eorec	asting		,	🔣 Or <u>d</u> ina	I		1
1	/		65	2	Supin	al			Probit.			1
1	0		1	2	Multin				B Nonlin	ear		1
2	9		-1	-1	wi <u>u</u> itip	ie Respons	ie .					2
2	1		53		💋 Missin	ng <u>V</u> alue Ar	alysis		weigh	t Estimation.		
2	2		49	1	Multip	le Imputati	on	>	2-Stag	e Least Squa	ares	1
		<			Comp	lex Sample	S	>	📠 Quanti	le		
Data	View	Variable	e View		📅 Simula	ation			🔠 Optima	al Scaling (C	ATREG)	
			Quality Control			>						

Figure 5.10: First Step When Choosing Logistic Regression

In our case, we have a dependent variable, "Enough knowledge about diabetes selfcare and management", with two categories, "yes, I think I have sufficient knowledge" and "no, I don't think I have enough knowledge', and five independent variables, gender, employment status, assessment of current health status, family history of diabetes, level of self-confidences. We use the "Enter" method that is commonly used for standard regression analysis (Fig. 5.11). You can choose from the following methods: Enter, Forward: Conditional, Forward: LR, Forward: Wald, Backward: Conditional, Backward: LR, Backward: Wald. The differences between the methods are in how they add or remove variables from the model. The enter method includes all predictor variables simultaneously without selection. In forward methods, the algorithm starts with an empty model and gradually adds predictors one by one until no more significant improvement is observed. Backward methods start with a full model including all possible predictors and gradually remove predictors one by one until no significant improvement is observed.



Figure 5.11: Choosing Variables When Conducting Logistic Regression

Next table contains the Cox & Snell R Squared and Nagelkerke R Squared values, which are both methods for calculating a variation (Fig. 5.12). The variance of the dependent variable in our model ranges from 22.1% to 32.4%.

		Chi-square	df	Sig.
Step 1	Step	14.483	5	.013
	Block	14.483	5	.013
	Model	14.483	5	.013

Omnibus Tests of Model Coefficients

Model Summary

Step	-2 Log	Cox & Snell R	Nagelkerke R
	likelihood	Square	Square
1	51.824 ^a	.221	.324

 Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Figure 5.12: Results (Model Summary) of Logistic Regression

The logistic regression estimates the probability that an event will happen, so in our case, whether a person will have enough knowledge about diabetes self-care and management.

Table index states, "The cut value is 0.500". This means that if the probability of a case being classified as "yes" is greater than 0.500, then that case is classified as "yes". Otherwise, the case is classified as "no" (as mentioned earlier).

In our case, the model classified the 39 people who thought they had sufficient knowledge about diabetes self-care and management into the appropriate group. It classified four people into the group with insufficient knowledge. When classifying persons with enough knowledge about diabetes self-care and management, the accuracy was 90.7%. 10 people who did not have enough knowledge were inappropriately classified as having enough knowledge. Five people were appropriately classified as not having enough knowledge. The model accuracy value was 33.3%. The overall model correctness was 75.9%.

Classification Table										
				Predicted						
			Do you think you knowledge abou care and ma	Do you think you have enough knowledge about diabetes self- care and management?						
	Observed		Yes, I think I have sufficient knowledge	, I think I No, I don't have think I have ufficient enough owledge knowledge						
Step 1	Do you think you have enough knowledge about	Yes, I think I have sufficient knowledge	39	4	90.7					
	diabetes self-care and management?	No, I don't think I have enough knowledge	10	5	33.3					
	Overall Percentage				75.9					

Classification Table^a

a. The cut value is .500

Figure 5.13: Results (Classification Table) of Logistic Regression

Figure 5.14 shows the contribution of each independent variable to the model and its statistical significance.

The Wald test (column " Wald ") is used to determine the statistical significance for each of the independent variables. In our case, none of the variables are statistically significant.

		В	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	Gender:	.677	.737	.846	1	.358	1.969
	Are you employed?	.063	.404	.025	1	.875	1.065
	Does anyone in your family have a diagnosis of diabetes?	126	.261	.233	1	.629	.881
	How would you rate your current health?	.621	.527	1.392	1	.238	1.861
	Confidence_descriptve	-1.551	.713	4.733	1	.030	.212
	Constant	1.881	3.317	.322	1	.571	6.561

Variables in the Equation

a. Variable(s) entered on step 1: Gender:, Are you employed?, Does anyone in your family have a diagnosis of diabetes?, How would you rate your current health?, Confidence_descriptve.

Figure 5.14: Results (Variables in the	Equation) of Logistic	Regression
i iguie 5.14. neouno	(anabies in the	L'Equation) of hogistic	regression

5.5 Multiple Regression

We speak of a multiple regression when a random variable depends on more than one explanatory variable. The variable we want to predict is called the dependent variable; the variables we are using to predict the value of the dependent variable are called the independent variables. The multiple regression also allows you to determine the overall fit (variance explained) of the model and the relative contribution of each of the predictors to the total variance explained.

Assumptions for performing multiple regression are:

- The dependent variable should be measured on a continuous scale (i.e., it is either an interval or ratio variable).
- Two or more independent variables must be included, which can be either continuous (i.e., an interval or ratio variable) or categorical (i.e., an ordinal or nominal variable).
- There should be independence of observations (i.e., independence of residuals),
 which we can easily check using the Durbin-Watson statistic.
- There needs to be a linear relationship between:
 - (a) the dependent variable and each of your independent variables,
 - (b) the dependent variable and the independent variables collectively.
- Data needs to show homoscedasticity, which is where the variances along the line of best fit remain similar as you move along the line.
- Data must not show multicollinearity, which occurs when you have two or more independent variables that are highly correlated with each other.
- There should be no significant outliers, high leverage points, or highly influential points.
- We need to check that the residuals (errors) are approximately normally distributed (we explain these terms in our enhanced multiple regression guide).

5.6 Steps to Conduct the Multiple Regression in SPSS

Click Analyze -> Regression -> Linear:

<u>Eile</u>	dit	View	Data	Transform	Analyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	Windo	w <u>H</u> elp		
					Po <u>w</u> e	r Analysis		>				Q
_		1994			Meta	Analysis		>				-
					Repor	Reports		>				
		I 🖓	21	💑 Q2	Desci	iptive Stati:	stics	>	6 Q9	💑 Q10	💑 Q11	💑 Q14
1			69	2	Baye	sian Statist	ics	>	4	2		1 1
2			45	2	Ta <u>b</u> le	s		>	4	3	:	2 1
3			59	1	Comp	are Means		>	-1	1		1 1
4			70	2	Gene	ral Linear N	lodel	>	4	2		1 1
5			43	2	Gana	ralized Line	ar Modele	,	4	1		1 1
6			83	2	Misse	Medele			3	3		3 2
7			45	2	INIZed	i wouers		í.	-1	2		1 -1
8			54	1	Corre	late		>	4	3		1 2
9			46	2	Regre	ssion		>	Auton	natic Linear N	lodeling	1
10			-1	1	Loglin	iear		>	🔣 Linea	r		2
11			54	1	Neura	I Networks		>	Curve	Estimation		1
12			46	1	Class	ify		>		I Least Squar		1
13			67	1	Dime	nsion Redu	ction	>	r aiua	ii Lea <u>s</u> t Oqual	es	1
14			-1	2	Scale			>	🔚 Binary	y Logistic		1
15			81	2	Nonp	arametric T	ests	>	La Multin	nomial Logistic	D	1
16			51	2	Eorec	asting		>	K Ordina	al		1
1/			65	2	Sunia	79		>	Probit			1
10			04	2	Multis	ui Ja Daanani		,	Nonlin	lear		
20			-1	-1	wiuitip	ne respons		,				2
20	_		53	2	Missir 🏭	ng <u>V</u> alue Ar	alysis		Ma vveigt	nt Estimation.		1
21			49	1	Multip	ole Imputati	on	>	10 2-Stag	ge Least Squa	ares	1
		<			Comp	lex Sample	s	>	🔠 Quant	tile		
Data V	/iew	Variab	le View		🖶 Simul	ation			🔠 Optim	nal Scaling (C	ATREG)	

Figure 5.15: First Step When Choosing Multiple Regression

Transfer the dependent variable (i.e., Confidences score) into the Dependent: box and the independent variables (i.e., age, gender, education, family history of diagnosed disease, marital status, assessment of current health status, sufficient knowledge of diabetes self-care and management) into the Independent(s): box:



Figure 5.16: Choosing Variables When Performing Multiple Regression

lick on the Statistics button:	
tinear Regression: Statisti	cs ×
Regression Coefficie Estimates Confidence intervals Level(%): 95 Covariance matrix	 ✓ Model fit □ R squared change □ Descriptives □ Part and partial correlations □ Collinearity diagnostics
Residuals D <u>u</u> rbin-Watson <u>C</u> asewise diagnostic <u>O</u> utliers outside: <u>A</u> II cases	s 3 standard deviations
Continue	Cancel Help

Figure 5.17: Statistics Options When Performing Multiple Regression (1)

In addition to the options that are selected by default, select Confidence intervals in the –Regression Coefficients– area, leaving the Level (%): option at "95":

the time ar Regression: Statist	ics		×
Regression Coefficie Estimates Confidence intervals Level(%): 95 Covariance matrix	✓ Mode R sq Desc Part Collin	el fit uared change riptives and partial corre nearity diagnost	elations
Residuals			
Durbin-Watson Casewise diagnostic Outliers outside:	3	standard devi	ations
<u>C</u> ontinue	Cancel	Help	

Figure 5.18: Statistics Options When Performing Multiple Regression (2)

Click on the button. You will be returned to the Linear Regression dialogue box. Click on the button. This will generate the output. The first table is the Model Summary table. This table provides the R, R2, adjusted R2, and the standard error of the estimate, which can be used to determine how well a regression model fits the data:

Model Summary								
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate				
1	.725 ^a	.525	.463	13.58999				

a. Predictors: (Constant), Do you think you have enough knowledge about diabetes self-care and management?, What is your marital status?, Education:, Does anyone in your family have a diagnosis of diabetes?, Gender:, Age (years):, How would you rate your current health?

Figure 5.19: Results (Model Summary) of Multiple Regression

The "R" column represents the value of R, the multiple correlation coefficient. R can be considered to be one measure of the quality of the prediction of the dependent variable, in this case, the confidence score. A value of 0.725, in this example, indicates a middle level of prediction. The "R Square" column represents the R2 value (also called the coefficient of determination), which is the proportion of variance in the dependent variable that can be explained by the independent variables (technically, it is the proportion of variation accounted for by the regression model above and beyond the mean model). You can see from our value of 0.525 that our independent variables explain 52.5% of the variability of our dependent variable, the confidence score.

The F-ratio in the ANOVA table (Fig. 5.20) tests whether the overall regression model is a good fit for the data. The table shows that the independent variables statistically significantly predict the dependent variable, F(7, 53) = 8.381, p<0.001.

	ANOVA ^a								
Мо	del	Sum of Squares	df	Mean Square	F	Sig.			
1	Regression	8954.810	7	1279.259	8.381	<.001 ^b			
	Residual	8089.628	53	152.634					
	Total	17044.438	60						

a. Dependent Variable: confidencesscore

b. Predictors: (Constant), Do you think you have enough knowledge about diabetes self-care and management?, What is your marital status?, Education:, Does anyone in your family have a diagnosis of diabetes?, Gender., Age (years):, How would you rate your current health?



The general form of the equation to predict blood sugar from age, gender, education, family history of the diagnosed disease, marital status, assessment of current health status, and sufficient knowledge of diabetes self-care and management is:

predicted confidences score = 111.437 + (0.067 x age) + (3.614 x gender) + (0.155 x education) + (0.124 x family history of diabetes) + (1.166 x marital status) - (7.878 x assessment of health status) - (20.753 x knowledge)

This is obtained from the Coefficients table (Fig. 5.21):

Model		Unstandardize B	d Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	111.437	14.396		7.741	<.001
	Age (years):	.067	.160	.045	.419	.677
	Gender:	3.614	3.416	.108	1.058	.295
	Education:	.155	1.164	.014	.133	.894
	Does anyone in your family have a diagnosis of diabetes?	.124	1.242	.010	.100	.921
	What is your marital status?	1.166	1.620	.074	.720	.475
	How would you rate your current health?	-7.878	2.452	353	-3.213	.002
	Do you think you have enough knowledge about diabetes self-care and management?	-20.753	4.612	494	-4.500	<.001

Coefficients^a

a. Dependent Variable: confidencesscore

Figure 5.21: Results (Coefficients) of Multiple Regression

Unstandardized coefficients indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant. Consider the effect of age in this example. The unstandardized coefficient, B1, for age, is equal to 0.067 (see Coefficients table). This means that for each one-year increase in age, there is a decrease in confidence score of 0.067.

You can test for the statistical significance of each of the independent variables. This tests whether the unstandardized (or standardized) coefficients are equal to 0 (zero) in the population. If p<0.05, you can conclude that the coefficients are statistically significantly different to 0 (zero). The t-value and corresponding p-value are in the "t" and "Sig." columns.



6 Questionnaire Validation

The chapter covers the steps of back translating a questionnaire, developing the questionnaire and validating the questionnaire. It also presents ways of assessing the reliability and validity of a questionnaire. The learning objectives of this chapter are to enable the reader to understand the whole process of questionnaire development, including back-translation techniques to ensure the linguistic and cultural appropriateness of the questionnaire. The reader will be introduced to the key steps in questionnaire design, such as determining the content, structure and format of the questions. In addition, the chapter will introduce the reader to validation methods, which include checking that the questionnaire measures what it is intended to measure. An important part of the chapter will also present ways of assessing the reliability and validity of the questionnaire.

6.1 Questionnaire Translation

Back-translation is necessary for the use of the questionnaire in the local environment. The original questionnaire should be translated into the target language by at least two independent translators. It is recommended that both translators are native speakers of the target language. This step is followed by a backtranslation. The translated questionnaire should be translated back into the source language by a second independent translator. The re-translated questionnaire in the original language is checked against the original questionnaire to identify any discrepancies. This step is crucial to ensure an accurate translation of the questionnaire (Tyupa, 2011; Behr, 2017).

6.2 Questionnaire Development

Steps in the process of questionnaire development (Tsang, et al., 2017):

- Identify the dimensionality of the construct.
- Determine the format in which the questionnaire will be administered.
- Determine the item format.
- Item development.
- Determine the intended length of the questionnaire.
- Review and revise the initial pool of items.
- Preliminary pilot testing.



Figure 6.1: Questionnaire Translation Steps

6.3 Reliability vs. Validity

Reliability is the extent to which the outcomes are consistent when the experiment is repeated more than once (Watson, 2015).

Internal consistency reflects the extent to which the questionnaire items are intercorrelated or whether they are consistent in measuring the same construct. Internal consistency is commonly estimated using the coefficient alpha - known as Cronbach's alpha. Cronbach's alpha ranges from 0 to 1 (when some items are negatively correlated with other items in the questionnaire, it is possible to have negative values of Cronbach's alpha).

6.4 Test-Retest Reliability

Test-retest reliability refers to the extent to which individuals' responses to the questionnaire items remain relatively consistent across repeated administration of the same questionnaire or alternate questionnaire forms. The correlation between the two questionnaires' responses can be referred to as the coefficient of stability. A larger stability coefficient indicates stronger test-retest reliability, reflecting that measurement error of the questionnaire is less likely to be attributable to changes in the individuals' responses over time (Yen & Lo, 2002).

6.5 Inter-Rater Reliability

For questionnaires in which multiple raters complete the same instrument for each examinee (e.g., a checklist of behaviour/symptoms), the extent to which raters are consistent in their observations across the same group of examinees can be evaluated. This consistency is referred to as inter-rater reliability or inter-rater agreement. Validity is the extent to which the instruments that are used in the experiment measure exactly what you want them to measure. The validity of a questionnaire is determined by analysing whether the questionnaire measures what it is intended to measure (Hallgren, 2012).

6.6 Content Validity

Content validity refers to the extent to which the items in a questionnaire are representative of the entire theoretical construct the questionnaire is designed to assess. The experts judge whether the questionnaire items are adequately measuring the construct intended to assess and whether the items are sufficient to measure the domain of interest. Several approaches to quantify the judgment of content validity across experts are also available, such as the content validity ratio and content validation form.



Figure 6.2: Example Items to Assess Content Validity

Figure 6.3 shows the division of CVI (content validity) into CVI for a single index and CVI for the whole scale (Polit, et al., 2007).



Figure 6.3: Content Validity Index

For the content checks, we prepare a content assessment form (Fig. 6.4) for the team of experts who will carry out the review. It is recommended to use the evaluation form (Fig. 6.5).

The selection of experts for the review is based on the individual's knowledge and expertise on the topic. Researchers recommend that the number of content reviewers should be at least six and no more than ten (Yusoff, 2019).

Dear Experts,
We will assess participants' self-care using the validated Self-Care of Diabetes Inventory (SCODI) (Ausili, et al., 2017).
We kindly ask you to rate each statement for content relevance on a 4-point scale:
 1 - content/statement not relevant/not understandable/not relevant;
2 - content / claim is poorly understood / very deficient / incomplete / partially relevant;
 3 - content / claim is partially understandable / partially relevant;
4 - the content / claim is fully understandable / fully relevant.
Please complete the table by placing an X under the selected rating and entering any comments, advice,
opinions on the content and translation in the "Comments" column.

Figure 6.4: Example of Guidance for Assessing the Relevance of Questions for Content Validation by Experts

SCODI C

Listed below are some of the behaviours a person with diabetes can do to help improve blood sugar levels when they are too high or too low. Participants are asked to identify how often they do (or would do) these actions when symptoms occur or when blood sugar is out of range (rated on a Likert scale from 1 to 5, where 1 represents never and 5 represents always; question 29 asks for a yes or no answer)?

		1	2	3	4	Comments
1.	Check your blood sugar when you feel symptoms (such as thirst, frequent urination, weakness, perspiration, anxiety)?					
2.	When you have abnormal blood sugar levels, do you take notes about the events that could have caused it and actions you took?					
з.	When you have abnormal blood sugar levels, do you ask a family member or friend for advice?					
4.	When you have symptoms, and you discover that your blood sugar is low, do you eat or drink something with sugar to solve the problem?					
5.	If you find out that your blood sugar is high, do you adjust your diet to fix it?					
6.	If you find out that your blood sugar is high, do you adjust your physical activity to fix it?					
7.	After taking actions to adjust an abnormal blood sugar level, do you re-check your blood sugar to assess if the actions you took were effective?					
8.	If you find out that your blood sugar is very low or very high, do you call your health care provider for advice?					
9.	Do you take insulin?					
10.	If you find out that your blood sugar is too high or too low, do you adjust your insulin dosage in the way your health care provider suggested?					

Figure 6.5: Example of a Form for Assessing the Relevance of Content by Experts

The results are then used to calculate I-CVI (item content validity index), S-CVI/Ave (scale-level content validity index based on the average method) and S-CVI/AU (scale-level content validity index based on the universal agreement method).

We scored 1 if the experts rated the items 3 (partially relevant) or 4 (fully relevant) and 0 if the experts rated them 1 (not relevant) or 2 (partially relevant).

Experts agree: represents the sum of all experts who agreed, i.e., gave a rating of 3 or 4. In our example, 5 experts agreed on Q1 (1+1+1+1+1+0=5)

I-CVI: Percentage of content experts who rated the relevance of the element as 3 or 4. In our case, I-CVI for Q21 is calculated as 5 divided by 6 experts and is 0.83.

<u>Formula</u>: I-CVI = (number of items rated 3 or 4)/(number of experts)

General agreement (UA): We score 1 for those questions where all experts agree. In our case, such questions were Q4, Q5, Q7 and Q9.

S-CVI/Ave: The average of the I-CVI scores for all items on the scale, or the average of the proportion of importance scored by all experts. The relevance score is the average of the relevance scores given by each expert.

<u>Formula</u>: S-CVI/Ave = (sum of I-CVI scores)/(number of items); S-CVI/Ave = (sum of relative importance ratings)/(number of experts)

S-CVI/UA: the proportion of items in the scale that achieve a relevance score of 3 or 4 across all experts. The universal agreement (UA) score is given as 1 if the item achieved 100% expert agreement. Otherwise, the UA score is given as 0.

Formula: S-CVI/UA = (sum of UA scores)/(number of items) (Yusoff, 2019).

No.	o. Question(s)		Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	N	Expert in Agreements	I-CVI	UA
1.	Check your blood sugar when you feel symptoms (such as thirst, frequent urination, weakness, perspiration, anxiety)?	1	1	1	1	1	0				
2.	When you have abnormal blood sugar levels, do you take notes about the events that could have caused it and actions you took?	1	1	1	1	1	0				
3.	When you have abnormal blood sugar levels, do you ask a family member or friend for advice?	1	1	1	1	1	1				
4.	When you have symptoms, and you discover that your blood sugar is low, do you eat or drink something with sugar to solve the problem?	1	1	1	1	1	0				
5.	If you find out that your blood sugar is high, do you adjust your diet to fix it?	1	1	1	1	1	0				
6.	If you find out that your blood sugar is high, do you adjust your physical activity to fix it?	1	1	1	1	1	1				
7.	After taking actions to adjust an abnormal blood sugar level, do you re-check your blood sugar to assess if the actions you took were effective?	1	1	1	1	1	0				
8.	If you find out that your blood sugar is very low or very high, do you call your health care provider for advice?	1	1	1	1	1	1				
9.	Do you take insulin?	1	1	1	1	1	0				
10.	If you find out that your blood sugar is too high or too low, do you adjust your insulin dosage in the way your health care provider suggested?	1	1	1	1	1	0				
									S-CVI/Ave		
-	Proportion relevance			5 (30/10)	o (based on r	reportion	(alayanca)		S-CVI/UA		
	S-CVI/Ave (based on proportion relevance)										

6.7 Construct Validity

Construct validity is the most important concept in evaluating a questionnaire that is designed to measure a construct that is not directly observable (e.g., pain, quality of recovery). If a questionnaire lacks construct validity, it will be difficult to interpret results from the questionnaire, and inferences cannot be drawn from questionnaire responses to a behaviour domain. The construct validity of a questionnaire can be evaluated by estimating its association with other variables (or measures of a construct) with which it should be correlated positively, negatively, or not at all (Smith, 2005; Bhandari, 2022).







7 Reliability Analysis

The chapter introduces reliability analysis, including intraclass correlation coefficient (ICC) and Cronbach's alpha, and the practical implementation of these analyses in SPSS software. The learning objectives of this chapter are to enable the reader to understand the key concepts and methods for assessing the reliability of survey questionnaires. The reader will learn how to use ICC to measure inter-rater reliability and Cronbach's alpha to assess the internal consistency of a questionnaire. The chapter also provides practical guidance on how to perform these analyses in SPSS.

The following sections consider, in turn: inter-rater reliability, test-retest reliability, and intra-rater reliability. These methods are all based on correlation between sets of measurement, and the test of choice for each is selected from one of a range of forms on intraclass correlations (ICCs) (Koo & Li, 2016). These have replaced other measures of reliability based on correlation, such as Pearson's r and the Kappa statistic, which were previously used. The reason that ICCs are preferred is that, as opposed to these other tests of correlation is that ICCs measure both the correlation between two sets of measurement (raters or ratings) and measure the degree of agreement between the two sets of measurement.

There are 10 different forms of ICC, but in the following, we will consider only the three that are most common and required to test these forms of reliability rigorously. ICCs are described in terms of 'model', 'type' and 'definition'. These depend,

respectively, on whether we are: 1. dealing with the same raters between ratings or a random sample of raters; 2. are we using the mean value of several raters or values from single raters; and 3. are we interested in consistency between raters or absolute agreement. For an excellent and detailed consideration of ICCs, consult Koo and Li (2016). As a rule of thumb, we are normally concerned with single-rater measurements. Where different raters are involved in rating the same phenomenon (inter-rater reliability), it is sufficient to look at consistency between measurements, but where the same raters are involved (test-retest reliability and intra-rater reliability), then we need to look at absolute agreement between ratings.

ICCs are normally expressed with 95% confidence intervals, and values, which lie between 0 - 1, are conventionally interpreted as follows:

Value	Explanation
< 0.05	Poor
0.05-0.75	Moderate
0.75-0.90	Good
>0.90	Excellent

Figure 7.1: Cronbach Alpha Value and Reliability Level

7.1 Intraclass Correlations

Click *Analyze* > *Scale* > *Reliability analysis*, as shown in **Figure 7.2**:

Eile Edit	View Data	Transform	Analyze Graphs Utilities	Extension	s <u>W</u> indo	w <u>H</u> elp		~	
			Meta Analysis	,		14 4	• 🖭 (L	
			Reports						
	🛷 Q1	💑 Q2	Descriptive Statistics	>	6 Q9	💑 Q10	💑 Q11	💑 Q14	SCODI_C
1	69	2	Bayesian Statistics	>	4	2	1	1	4
2	45	2	Tables	>	4	3	2	1	5
3	59	1	Compare Means	>	-1	1	1	1	5
4	70	2	General Linear Model	>	4	2	1	1	-3
5	43	2	Generalized Linear Models	,	4	1	1	1	3
6	83	2	Mined Medele		3	3	3	2	3
7	45	2	Mixed Woders	<i>(</i>	-1	2	1	-1	-3
8	54	1	Correlate	>	4	3	1	2	-3
9	46	2	Regression	>	1	2	1	1	5
10	-1	1	Loglinear	>	1	2	1	2	3
11	54	1	Neural Networks	>	4	3	1	1	5
12	46	1	Classify	>	1	2	1	1	3
13	67	1	Dimension Reduction	>	1	2	2	1	3
14	-1	2	Casla		-	2	2		5
15	81	2	OCAIR		Reliat	ility Analysis.			5
16	51	2	Nonparametric Tests	,	📶 Weigh	nted Kappa			1
17	65	2	Forecasting	>	Multid	imensional Ur	folding (PREF	SCAL)	3
18	64	2	Survival	>	Multid	imensional Sc	aling (PROXS	CALL	5
19	-1	-1	Multiple Response	>					4
20	59	2	Missing Value Analysis		Multid	imensional Sc	aling (ALSCA	L)	1

Figure 7.2: First Step When Conducting Reliability Level

Transfer the selected variables. Leave the model Alpha, as this is the command to run Cronbach's Alpha. Select the options highlighted in **Figure 7.3**.

			🍓 Reliability Analysis: Statistics	×
			Descriptives for	Inter-Item
			[tem]	Correlations
			Scale	Covariances
🍓 Reliability Analysis		×	Scale if item deleted	
	Items:	Statistics	Summaries	ANOVA Table
Age (years): [Q1]	💰 Check your blood sugar when y 📤		Means	
🗞 Gender: [Q2]	💫 When you have abnormal blood		∐ ⊻ariances	OEtest
& Education: [Q3]	When you have abnormal blood		Covariances	O Friedman chi-sguare
Are you employed? [Q4]	Vvnen you nave symptoms, and		Correlations	O Cochran chi-square
Does anvone in your family hav	A If you find out that your blood su			
Row would you rate your current	After taking actions to adjust an		Interrater Agreement: Fleiss Kappa	
💫 How would you rate your commi	A		Display agreement on Individual	categories
Do you think you have enough k	Ratings:		Ignore string cases	
Check your blood sugar when y			String category labels are di	splayed in uppercase
When you have abnormal blood			Asymptotic significance level (%):	95
When you have symptoms and			Missing	
& If you find out that your blood su			Exclude <u>b</u> oth user-missing and s	system missing values
lf you find out that your blood su			O User-missing values are treated	as valid
After taking actions to adjust an			Hotelling's T-square	Tukey's test of additivity
De you take insulia? [SCODI _00]			Intraclass correlation coefficient	
Model: Alpha v			Mo <u>d</u> el: Two-Way Mixed	Type: Consistency
Scale label:			Confidence interval: 95 %	Test val <u>u</u> e: 0
OK Paste	e <u>R</u> eset Cancel Help		Continue Can	cel Help

Figure 7.3: Choosing Variables When Performing Reliability Analysis

The first table of results presents the reliability, which is shown by the actual value for Cronbach's alpha. For our questionnaire, the Cronbach alpha is 0.780, which represents good reliability.

Reliability Statistics							
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items					
.780 .694 10							

Figure 7.4: Reliability Statistic Analysis (Cronbach's Alpha)

The most important column in the table is "Cronbach's Alpha if Item Deleted". This column represents the value of Cronbach's alpha if this item were to be deleted (**Fig. 7.5**).

Item-Total Statistics														
	Scale Corrected Squared Cronbach's Scale Mean if Variance if Item-Total Multiple Alpha if Item Item Deleted Item Deleted Correlation Correlation Deleted													
Check your blood sugar when you feel symptoms (such as thirst, frequent urination, weakness, perspiration, anxiety)?	31.12	32.760	.575	.617	.747									
When you have abnormal blood sugar levels, do you take notes about the events that could have caused it and actions you took?	31.73	30.101	.634	.616	.734									
When you have abnormal blood sugar levels, do you ask a family member or friend for advice?	32.59	33.999	.281	.272	.790									
When you have symptoms, and you discover that your blood sugar is low, do you eat or drink something with sugar to solve the problem?	30.93	31.020	.619	.616	.738									
If you find out that your blood sugar is high, do you adjust your diet to fix it?	31.17	31.295	.686	.696	.732									
If you find out that your blood sugar is high, do you adjust your physical activity to fix it?	31.59	32.799	.543	.634	.750									
After taking actions to adjust an abnormal blood sugar level, do you re- check your blood sugar to assess if the actions you took were effective?	30.88	30.110	.844	.814	.713									
If you find out that your blood sugar is very low or very high, do you call your health care provider for advice?	33.44	39.702	051	.348	.826									
Do you take insulin?	34.34	42.730	697	.774	.810									
If you find out that your blood sugar is too high or too low, do you adjust your insulin dosage in the way your health care provider suggested?	30.95	30.698	.617	.818	.738									



7.2 Inter-Rater Reliability

For questionnaires in which multiple raters complete the same instrument for each examinee (e.g., a checklist of behaviour/symptoms), the extent to which raters are consistent in their observations across the same group of examinees can be evaluated. This consistency is referred to as inter-rater reliability or inter-rater agreement. Normally, we are concerned with being able to extrapolate our results to the general population from a randomly selected group of raters, and the appropriate

form of ICC to test this is a Two-way random effects model with consistency based on single raters. If we were only specifically interested in the particular group of raters participating in our study, we would use a mixed-effects model.

Click *Analyze* > *Scale* > *Reliability analysis*, as shown in **Figure 7.6**:

tate 🕼	r.sav [Da	ataSet2] -	IBM SP	SS Statistics Data	Editor												
<u>F</u> ile	<u>E</u> dit	View	<u>D</u> ata	Transform	Analyze	<u>G</u> raphs	<u>U</u> tilities	Extensions	<u>W</u> indov	v <u>H</u> e	lp						
				K 3	Po <u>w</u> er	Analysis		>			\bigcirc	•	Q				
22 · D	ator6				Meta	Analysis		>									-
22 . R	atero				Repor	ts		>			_						_
		💑 Ra	ter1	Kater2	Descr	iptive Stati	stics	>	Rater6	var		var	var		var	var	
	1		4.00	4.00	Baves	ian Statist	ics	>	2.00								
	2		4.00	3.00	Tabler			,	1.00								
	3		3.00	4.00	i agies	·			3.00								
4	4		4.00	4.00	Comp	are Means		>	2.00								
	5		4.00	4.00	Gener	al Linear N	lodel	>	2.00								
	6		4.00	3.00	Gener	ali <u>z</u> ed Line	ar Models	>	3.00								
1	7		3.00	4.00	Mixed	Models		>	2.00								
- 1	3		4.00	4.00	Correl	ate		>	3.00								
	9		4.00	4.00	Dogro	eeion		>	1.00								
1	0		4.00	3.00	Kegre	331011		,	2.00								
1	1		4.00	3.00	Loglin	ear		,	2.00								
1	2		4.00	4.00	Neura	I Networks		>	2.00								
1	3		4.00	4.00	Class	fy		>	3.00								
1	4		4.00	4.00	Dimer	ision Redu	ction	>	2.00								
1	5		4.00	4.00	Sc <u>a</u> le			>	Reliab	ility An:	alvsis						
1	6		4.00	4.00	Nonna	rametric T	ests	>			aryoro						
1	7		4.00	4.00	Eoroo	notina	0010		VVeigh	ited <u>K</u> ap	ppa						
1	8		4.00	4.00	TOIEC	as <u>t</u> ing			🛃 M <u>u</u> ltidi	imensio	onal Unf	folding (l	PREFSCA	L)			
1	9		4.00	4.00	Surviv	81		>	👪 Multidi	imensio	nal Sca	aling (<u>P</u> F	ROXSCAL)				
2	0		4.00	4.00	M <u>u</u> ltip	le Respons	se	>	R Multidi	imensio	nal Sca	aling (Al	SCAL)				
2	1		4.00	4.00	월 Missin	g <u>V</u> alue Ar	nalysis		2.00								
2	2				Multin	le Imnutati	on	>									

Figure 7.6: First Step When Conducting Reliability Level

			Neliability Analysis: Statistics	×
			Descriptives for tem Scale Scale if item deleted	Inter-Item Correlations Covarianc <u>e</u> s
ta Reliability Analysis	terns:	× Statistics	Summaries Means Variances Covariances	ANOVA Table None E test Friedman chi-square Cocheng abia genera
éo, Rater3 éò, Rater4 éò, Rater5 éò, Rater5	Kater3 Ater4 Ater5 Rater5 Rater5		Correlations Interrater Agreement: Fleiss' Kappa Display agreement on individual Ignore string cases String category labels are di Asymptotic significance level (%):	categories
			Missing © Exclude both user-missing and to O User-missing values are treated Hotelling's T-square [Intraclass correlation coefficient]	system missing values as valid Tukey's test of additivity
Model: Alpha ~ Scale label:	a Read Cased Hala]	Mogel: Two-Way Mixed <u>C</u> onfidence interval: 95 % Continue Can	Type: Consistency ~ Test value: 0

Figure 7.7: Reliability Analysis: Statistics

Figure 7.8 shows the results of the reliability analysis.

Reliability S	tatistics
Cronbach's Alpha ^a	N of Items
184	6
a. The value i negative du negative av covariance items. This reliability m assumptio may want t item coding	s verage among s violates nodel ns. You o check gs.

Intraclass Correlation Coefficient

	Intraclass	95% Confide	ence Interval	F Test with True Value 0				
	Correlation	Lower Bound Upper Bound		Value	df1	df2	Sig	
Single Measures	027 ^a	100	.122	.844	20	100	.655	
Average Measures	184°	-1.189	.454	.844	20	100	.655	

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Figure 7.8: Results of the Reliability Analysis

7.3 Test-Retest Reliability

Test-retest reliability refers to how individuals' responses to the questionnaire items remain relatively consistent across repeated administration of the same questionnaire or alternate questionnaire forms. The appropriate form of ICC to use here is a Two-way mixed effects model with absolute agreement based on single raters.

ile <u>E</u> dit	View	Data	Transform	Analyze	Graphs	Utilities	Extensions	Window	w <u>H</u> elp				
			100	Power	Analysis		>			à 🗖 🤇	2		
2 : Rater6				Meta A	Analysis s		>	0	1.4				
	💰 Rat	er1	Rater2	Descri	ptive Stati	istics	>	Rater6	var	var	var	var	var
1		4.00	4.00	Rause	ino Statici	tice	,	2.00					
2		4.00	3.00	Dages	ian Statis	uco		1.00					
3		3.00	4.00	Tagies			'	3.00					
4		4.00	4.00	Compa	are Means		>	2.00					
5		4.00	4.00	Genera	al Linear N	Nodel	>	2.00					
6		4.00	3.00	Genera	alized Line	ear Models	>	3.00					
7		3.00	4.00	Mixed	Models		>	2.00					
8		4.00	4.00	Corrol	to			3.00					
9		4.00	4.00	Zonen			2	1.00					
10		4.00	3.00	Rediet	ssion		,	2.00					
11		4.00	3.00	Logline	par		,	2.00					
12		4.00	4.00	Neural	Networks		>	2.00					
13		4.00	4.00	Classi	fy		>	3.00					
14		4.00	4.00	Dimen	sion Redu	iction	>	2.00					
15		4.00	4.00	Scale			>	D. D. Lab	Star Arrebo				
16		4.00	4.00	Manaa	remetrie T	Tente .		Kenap	mity Analys	AS			
17		4.00	4.00	Nonpa	rannethic I	0313		10 Weigh	ited Kappa				
18		4.00	4.00	Foreca	isting		,	🔛 Multid	imensional	Unfolding (PF	REFSCAL)		
19		4.00	4.00	Surviva	el .		>	Multid	imensional	Scaling (PRC	XSCAL)		
20		4.00	4.00	Multipl	e Respon	se	>	EN Multid	imancional	Scaling (ALS	CALL		
21		4.00	4.00	Missin	y Value A	nalysis		SIS HOUD	mensional	Scamg (ALS	onij		
22	1			Multiel	e Imoutati	ion	,						

Figure 7.9: First Step When Conducting the Test-Retest Reliability

		t.	Reliability Analysis: Statistics	×
			Descriptives for	Inter-Item
			tem [tem	Correlations
			<u>S</u> cale	Covarianc <u>e</u> s
			Scale if item deleted	
			Summaries	ANOVA Table
			Means	
			<u>Variances</u>	O <u>F</u> test
			Covariances	O Friedman chi-sguare
ta Reliability Analysis		×	Correlations	O Cochran chi-square
& Rater1	Items:	Statistics	Interrater Agreement: Fleiss' Kappa	
& Rater2	Rater2		Display agreement on individual ca	ategories
Rater3 Rater4	S Rater3		Ignore string cases	
S Rater5	Rater5		String category labels are disp	played in uppercase
			Asymptotic significance level (%): 95	5
	Ratings:		Missing	
			\odot Exclude <u>b</u> oth user-missing and sy	stem missing values
			O User-missing values are treated as	s valid
•			Hotelling's T-square	Tukey's test of additivity
			Intraclass correlation coefficient	
			Mo <u>d</u> el: Two-Way Mixed 🗸	Type: Absolute Agreement ~
Model: Alpha V			Confidence interval: 95 %	Test val <u>u</u> e: 0
OK Paste	Reset Cancel Help		<u>C</u> ontinue Cance	l Help

Figure 7.10: Test-Retest Reliability

Figure 7.11 shows the results of the test-retest reliability analysis.



Intraclass Correlation Coefficient

	Intraclass	95% Confide	ence Interval	F Test with True Value 0				
	Correlation	Lower Bound	Upper Bound	Value	df1	df2	Sig	
Single Measures	009 ^a	035	.050	.844	20	100	.655	
Average Measures	060°	255	.240	.844	20	100	.655	

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Figure 7.11: Result of the Test-Retest Reliability

7.4 Intra-Rater Reliability

Intra-rate reliability refers to the extent to which individuals rate the same phenomenon and remain relatively consistent across repeated ratings. As with testretest reliability, the appropriate form of ICC to use here is a Two-way mixed effects model with absolute agreement based on single raters.

ta *Un	titled1 [[DataSet0] - I	BM SPS	S Statistics Da	ta Editor												
Eile	Edit	View D	lata	Transform	Analyze	Graphs	Utilities	Extensions	s <u>)</u>	Mindow	Help						
				53	Po <u>w</u> e	r <mark>Analys</mark> is		>		.		0	. (2			
10 :					Meta	Analysis rts		> >					_	•			
	1 2 3 4 5 5 5 7 7 3 9 0 1 1 2 2 3 4	& Confi	dent2 77.27 93.18 79.55 68.18 72.73 70.45 61.36 72.73	Conf Conf	Desci Desci Baye: Table: Comp <u>G</u> ener Gener Mixeo <u>C</u> orrel <u>R</u> egre Loglin Neura Class Dimer	riptive Stati sian Statis s pare Means ral Linear M ralized Line d Models late ession ear al Networks ify nsion Redu	istics s Vodel ear Models s sction	> > > > > > > > > > > > > > > > > > >		var		var	Va	r	var	Var	var
1	5		97.73	3	Scale		iononi	>		Reliabilit	y Anal	vsis					
1	6 7 8		68.18 72.73 100.00	3 3 0	<u>N</u> onpa Forec Surviv	arametric T as <u>t</u> ing al	lests .	>	K	Weighter M <u>u</u> ltidim	d <u>K</u> app ension	a al Unfol	ding (PR	EFSC	۹L)		
2	9 0 1		22.73	- 3 0	M <u>u</u> ltip Missir	ole Respon ng <u>V</u> alue A	se nalysis	>		Multidim <u>M</u> ultidim	ension ension	al Scali al Scali	ng (<u>P</u> RO ng (ALSO	XSCAL	.)		
4	6		50.5		Wultip	ne imputat	ion	,									

Figure 7.12: First Step When Conducting the Intra-Rater Reliability

			halve a contraction and the set of the set o	×
			Descriptives for tem Scale Scale if item deleted	Inter-Item Correļations Covarianc <u>e</u> s
 Reliability Analysis Confident2 Confident1 	Items:	×	Summaries Means Variances Covariances Correlations	ANOVA Table None E test O Friedman chi-sguare O Coc <u>h</u> ran chi-square
	Ratings:		Interrater Agreement: Fleiss' Kappa Display agreement on individual Ignore string cases String category labels are di Asymptotic significance level (%):	categories splayed in uppercase 95
	•		Missing © Exclude <u>b</u> oth user-missing and s O User-missing values are treated a Hotelling's T-square Ø Intraclass correlation coefficient	system missing values as valid ☐ Tu <u>k</u> ey's test of additivity
Model: Alpha Alpha	Paste Reset Cancel Help		Mo <u>d</u> el: Two-Way Mixed <u>C</u> onfidence interval: 95 % <u>Continue</u> Can	Type: Absolute Agreement Test value: Help

Figure 7.13: Intra-Rater Reliability

Case Processing Summary

		Ν	%
Cases	Valid	116	82.3
	Excluded ^a	25	17.7
	Total	141	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items		
1.000	2		

Intraclass Correlation Coefficient

	Intraclass	95% Confidence Interval		F Test with True Value 0			
	Correlation®	Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.957 ^a	.021	.991		115		
Average Measures	.978°	.042	.996		115		

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Figure 7.14: Result of the Intra-Rater Reliability



8 Factor Analysis

This chapter introduces factor analysis, including a detailed explanation of the basic concepts, the steps to perform principal component analysis (PCA) in SPSS software, and the interpretation of the results. The learning objectives of this chapter are to enable the reader to understand the concepts and methods associated with factor analysis, particularly PCA. The reader will learn about the purpose and usefulness of factor analysis in exploring data and will gain practical skills in performing PCA in SPSS and interpreting the results.

Factor analysis aims to reduce the dimensionality of the original space and to give an interpretation of the new space, which includes a reduced number of new dimensions that are supposed to be the basis for the old ones (Rietvel & van Hout, 1993, p. 254). It is used in areas such as medicine and nursing, economics, behavioural and social sciences, and geography (Yong & Pearce, 2013). Different methods of factor analysis are presented in Figure 8.1.



Figure 8.1: Types of Factor Analyses

There are a few methods to verify the adequacy of factor analysis (Nakazawa, 2011):

- criteria for the adequacy of the sample size,
- Kaiser-Meyer-Olkin (KMO) sampling adequacy criteria, which test whether a large number of factors exist in a data set,
- Bartlett's sphericity test, which tests the hypothesis that the correlations between the variables are greater than would be accidentally expected.

Commonly used analysis is Principal Components Analysis (PCA). It is a multivariate data reduction technique that aims to find a new set of variables that is equal in number to the original set of variables, with these synthetic variables not related (Rossiter, 2017). PCA is used to extract maximum variance from the data set

with each component to reduce a large number of variables into a smaller number of their weighted sums (components). PCA is conceptually different from factor analysis but is often used synonymously in practice.

PCA can be used for different purposes:

- If we have a large number of variables (questions/statements in a questionnaire) and we believe that some of the variables measure the same phenomenon/construct; if we find a high correlation between the variables, we can include only those variables that best represent the construct, and remove the rest.
- This method can be used when we want to create a new questionnaire/measurement scale, but we are not sure that all the variables (questions/statements) we have included in our tool measure the same construct. This way, we can make sure that the variables we have included are sufficiently representative or that we need to remove them.
- In the case where we want to include as few variables (questions/statements) as possible in an existing questionnaire/measurement scale to shorten it.

To carry out a PCA, the following assumptions must be met:

- Our data must include several variables that must be measured continuously (ordinal variables are often used).
- There must be a linear relationship between all variables.
- A sufficiently large sample size is needed to ensure the reliability of the results obtained. As a general rule, a minimum of 150 cases is recommended, or 5 to 10 cases per variable.
- The data must be suitable for data reduction, meaning there must be an adequate correlation between the variables for them to be reduced.
- There must be no significant outliers in the variables.

8.1 Steps to Conduct PCA in SPSS

We eliminate n components and decide on a number. Criteria (by relevance):

1. substantive meaningfulness,

- 2. scree plot,
- 3. Desired eigenvalue> 1.

If more than one component is eliminated, as a rule, we rotate - repeat the analysis with the appropriate number of eliminated components.

Orthogonal rotations (e.g., Varimax): uncorrelated components, fewer inappropriate solutions.

Oblique-angle rotations (e.g., direct oblimin): a simpler interpretation. If the correlations between the components are low (e.g., <0.1), we prefer to use Varimax.

We interpret the rotated components in terms of content ("what variables with high absolute saturations" have in common).

We calculate component achievements if desired:

- Descriptives: Univariate descriptives, correlations, reproduced correlations, etc.
- Important: At least some correlations should be at least medium-high.
- **Extraction:** The process of extracting components.
- Method: Principal components.
- **Extract:** The number of components extracted (in the non-rotated solution anyway you can start n).
- **Display:** Scree-plot (number of "large" components).
- Analyse: If variables with a higher variance are to have a greater effect on the results, we analyse the covariance matrix (usually for item analysis), otherwise correlatively (usually for test analysis).
- Rotation: Rotation selection.
- Loading plot: Draws a pattern and not correlations (structure) when rotating it at an angle.
- Scores: Save as variables: calculating people's achievements.
- Display factor score coefficient matrix: displays weights for calculating components.

In the following text, the process of conducting PCA in SPSS is presented in figures. To perform the PCA we should click the following: Click *Analyze -> Dimension Reduction -> Factor.*


Figure 8.2: First Step when Choosing PCA

Select the variables and specify the appropriate settings, as shown in Figure 8.3.

🚰 Factor Analysis	×	🗣 Factor Analysis: Descriptives 🛛 🕹 👋
✓ Age (years): [Q1] ^ Gender: [Q2] Action: [Q3] Are you employ	high or I Advice a uur medi in follow	Statistics
	your blo and if y ize the v ariable: ancel Help	Correlation Matrix Coefficients Inverse Significance levels Reproduced Determinant Anti-image KMO and Bartlett's test of sphericity Continue Cancel
Factor Analysis: Extraction ×	Factor Analysis: Rotation	× 🏰 Factor Analysis: Options >
Analyze Display © Cognelation matrix © Cogniance matrix Extract	○ None ○ Quartimax ○ Varimax ○ Equamax ○ Direct Qblimin ○ Promax	Missing Values O Exclude cases listwise Exclude cases pairwise O Replace with mean
Based on Eigenvalue Eigenvalues greater than: Fixed number of factors Factors to extract:	Display Rotated solution Loading	Coefficient Display Format ✓ Sorted by size plot(s) ✓ Suppress small coefficients
Magimum Iterations for Convergence: 25	Maximum Iterations for Converger	ace: 25 Absolute value below: 30 Absolute value below: 40 Absolute value

Figure 8.3: Choosing the Right FA Method

This table shows two tests that indicate the suitability of the data for FA implementation. Kaiser-Meyer-Olkin is a statistic that shows the proportion of variance in variables that can be caused by underlying factors. High values (close to 1.0) generally indicate that FA is beneficial. If the value is less than 0.50, FA results are unlikely to be useful. Bartlett's test of sphericity tests the hypothesis that variables are unrelated and, therefore, unsuitable for FA. Small values (less than 0.05) of significance level indicate that FA is beneficial.

KMO and Bartlett's Test					
Kaiser-Meyer-Olkin Measure	.890				
Bartlett's Test of Sphericity	Approx. Chi-Square	855.620			
	df	55			
	Sig.	<.001			

Figure 8.4: KMO and Bartlett's Test

Eigenvalue presents the quality of the result; a higher result means higher quality.

otal Variance	Explained
---------------	-----------

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.337	57.612	57.612	6.337	57.612	57.612	4.102	37.295	37.295
2	1.033	9.386	66.999	1.033	9.386	66.999	3.267	29.704	66.999
3	.811	7.373	74.372						
4	.626	5.688	80.060						
5	.545	4.951	85.010						
6	.401	3.643	88.654						
7	.361	3.285	91.939						
8	.295	2.682	94.621						
9	.272	2.471	97.092						
10	.175	1.595	98.686						
11	.144	1.314	100.000						

Extraction Method: Principal Component Analysis.

Figure 8.5: Table of Components and Eigenvalues

		Prevent high or low blood sugar levels and its symptoms.	Follow advice about nutrition and physical activity.	Take your medicines in the appropriate way (including insulin if prescribed).	Persist in following the treatment plan even when it's difficult.	Monitor your blood sugar as often as your health care provider asked you to.	Understand if your blood sugar levels are good or not.	Recognize the symptoms of low blood sugar.	Persist in monitoring your diabetes even when it s difficult	Take action to adjust your blood sugar and relieve your symptoms.	Evaluate if your actions were effective to change your blood sugar and relieve your symptoms.	Persist in carrying out actions to improve your blood sugar even when it s difficult.
Correlation	Prevent high or low blood sugar levels and its symptoms.	1.000	.648	.416	.507	.321	.455	.423	.481	.558	.684	.559
	Follow advice about nutrition and physical activity.	.648	1.000	.402	.534	.383	.345	.428	.503	.563	.640	.605
	Take your medicines in the appropriate way (including insulin if prescribed).	.416	.402	1.000	.592	.362	.428	.347	.479	.434	.439	.384
	Persist in following the treatment plan even when it's difficult.	.507	.534	.592	1.000	.381	.391	.452	.640	.527	.633	.626
	Monitor your blood sugar as often as your health care provider asked you to.	.321	.383	.362	.381	1.000	.641	.413	.571	.569	.480	.473
	Understand if your blood sugar levels are good or not.	.455	.345	.428	.391	.641	1.000	.525	.601	.585	.460	.470
	Recognize the symptoms of low blood sugar.	.423	.428	.347	.452	.413	.525	1.000	.614	.708	.592	.499
	Persist in monitoring your diabetes even when it's difficult.	.481	.503	.479	.640	.571	.601	.614	1.000	.730	.620	.635
	Take action to adjust your blood sugar and relieve your symptoms.	.558	.563	.434	.527	.569	.585	.708	.730	1.000	.736	.736
	Evaluate if your actions were effective to change your blood sugar and relieve your symptoms.	.684	.640	.439	.633	.480	.460	.592	.620	.736	1.000	.794
	Persist in carrying out actions to improve your blood sugar even when it' s difficult.	.559	.605	.384	.626	.473	.470	.499	.635	.736	.794	1.000
Sig. (1-tailed)	Prevent high or low blood sugar levels and its symptoms.		<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
	Follow advice about nutrition and physical activity.	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
	Take your medicines in the appropriate way (including insulin if prescribed).	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	Persist in following the treatment plan even when it's difficult.	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	Monitor your blood sugar as often as your health care provider asked you to.	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
	Understand if your blood sugar levels are good or not.	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
	Recognize the symptoms of low blood sugar.	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
	Persist in monitoring your diabetes even when it's difficult.	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
	Take action to adjust your blood sugar and relieve your symptoms.	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
	Evaluate if your actions were effective to change your blood sugar and relieve your symptoms.	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
	Persist in carrying out actions to improve your blood sugar even when it' s difficult.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

Correlation Matrix

Figure 8.6: Item Correlations

Utilities can range from 0 to 1. Higher values mean that a larger proportion of the variance of a particular variable can be explained by other variables. These are more reliable (Fig. 8.7).

	Initial	Extraction
Prevent high or low blood sugar levels and its symptoms.	1.000	.661
Follow advice about nutrition and physical activity.	1.000	.686
Take your medicines in the appropriate way (including insulin if prescribed).	1.000	.386
Persist in following the treatment plan even when it's difficult.	1.000	.628
Monitor your blood sugar as often as your health care provider asked you to.	1.000	.700
Understand if your blood sugar levels are good or not.	1.000	.744
Recognize the symptoms of low blood sugar.	1.000	.567
Persist in monitoring your diabetes even when it's difficult.	1.000	.726
Take action to adjust your blood sugar and relieve your symptoms.	1.000	.767
Evaluate if your actions were effective to change your blood sugar and relieve your symptoms.	1.000	.792
Persist in carrying out actions to improve your blood sugar even when it' s difficult.	1.000	.712

Communalities

Extraction Method: Principal Component Analysis.

Figure 8.7: The Utility Matrix

The rotated component matrix tells us which variables are most strongly associated with each component.

Rotated Component Matrix^a

	Component		
	1	2	
Follow advice about nutrition and physical activity.	.810		
Evaluate if your actions were effective to change your blood sugar and relieve your symptoms.	.802	.387	
Prevent high or low blood sugar levels and its symptoms.	.789		
Persist in carrying out actions to improve your blood sugar even when it' s difficult.	.738	.409	
Persist in following the treatment plan even when it's difficult.	.731	.307	
Take your medicines in the appropriate way (including insulin if prescribed).	.508	.359	
Understand if your blood sugar levels are good or not.		.837	
Monitor your blood sugar as often as your health care provider asked you to.		.818	
Persist in monitoring your diabetes even when it's difficult.	.508	.684	
Take action to adjust your blood sugar and relieve your symptoms.	.583	.654	
Recognize the symptoms of low blood sugar.	.416	.628	

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Figure 8.8: Component and Rotated Component Matrix



References

- Andrade, C., 2020. Sample size and its importance in research. Indian journal of psychological medicine, 42(1), pp.102-103.
- Ausili, D., Barbaranelli, C., Rossi, E., Rebora, P., Fabrizi, D., Coghi, C., Luciani, M., Vellone, E., Di Mauro, S. and Riegel, B., 2017. Development and psychometric testing of a theory-based tool to measure self-care in diabetes patients: the Self-Care of Diabetes Inventory. BMC Endocrine Disorders, 17(1), pp.1-12.
- Barton, B. and Peat, J., 2014. Medical statistics: A guide to SPSS, data analysis and critical appraisal. John Wiley & Sons.
- Behr, D., 2017. Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology*, 20(6), pp.573-584.
- Bhandari, P., 2022. Construct Validity | Definition, Types, & Examples. (Online) Available at: https://www.scribbr.com/methodology/construct-validity/ [Accessed 13 March 2019].
- Dugard, P. and Todman, J., 1995. Analysis of pre-test-post-test control group designs in educational research. *Educational Psychology*, 15(2), pp.181-198.
- Gelman, A. and Jennifer H., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.
- Goertzen, M.J., 2017. Introduction to quantitative research and data. *Library Technology Reports*, 53(4), pp.12-18.
- Hallgren, K.A., 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), p.23.
- Hastie et al., 2009. The Elements of Statistical Learning: Data mining, inference, and prediction, Springer.
- Holton, E.F. and Burnett, M.F., 2005. The basics of quantitative research. Research in organizations: Foundations and methods of inquiry, pp.29-44.
- Hoy, W.K. and Adams, C.M., 2015. Quantitative research in education: A primer. Sage Publications.
- James et al., 2013. An Introduction to Statistical Learning: With applications in R, Springer.
- Kim, H.Y., 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restorative dentistry & endodontics, 42(2), pp.152-155.
- Koo, T.K. and Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), pp.155-163.
- Liguori, G.R. and Moreira, L.F.P., 2018. Operating with Data-Statistics for the cardiovascular surgeon: Part I. Fundamentals of Biostatistics. *Brazilian journal of cardiovascular surgery*, 33, pp.III-VIII.
- Marsden, E. and Torgerson, C.J., 2012. Single group, pre-and post-test research designs: Some methodological concerns. Oxford Review of Education, 38(5), pp.583-616.
- Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. and Keshri, A., 2019. Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, 22(1), p.67.
- Mishra, P., Singh, U., Pandey, C.M., Mishra, P. and Pandey, G., 2019. Application of student's t-test, analysis of variance, and covariance. *Annals of cardiac anaesthesia*, 22(4), p.407.
- Nakazawa, M., 2011. R practice: Factor analysis. (Pdf) Available at: http://minato.sip21c.org/swtips/factor-in-R.pdf [Accessed 13 March 2019].
- Polit, D.F., Beck, T. and Owen, S.V., 2007. Focus on research methods is the CVI an acceptable indicator of content validity. *Research in Nursing & Health*, 30(4), pp.459-67.

Reaves, C.C., 1992. Quantitative research for the behavioral sciences. John Wiley & Sons.

- Rietvel, T. & van Hout, R., 1993. Statistical Techniques for the Study of Language and Language Behaviour. Berlin: Mouton de Gruyter.
- Rossiter, D. G., 2017. Tutorial: An example of statistical data analysis using the R environment for statistical computing. (Pdf) Available at: http://www.css.cornell.edu/faculty/dgr2/teach/R/R_corregr.pdf [Accessed 15 March 2019].
- Simkus, J., 2022, Simple Random Sampling: Definition, Steps and Examples. Simply Psychology. (Online) Available at: www.simplypsychology.org/simple-random-sampling.html_[Accessed 13 March 2019].
- Simkus, J., 2022, Jan 30. Convenience Sampling: Definition, Method and Examples. Simply Psychology. (Online) Available at: www.simplypsychology.org/convenience-samplinghtml [Accessed 13 March 2019].
- Smith, G.T., 2005. On construct validity: issues of method and measurement. *Psychological assessment*, 17(4), p.396.
- Sperandei, S., 2014. Understanding logistic regression analysis. Biochemia medica, 24(1), pp.12-18.
- Spieth, P.M., Kubasch, A.S., Penzlin, A.I., Illigens, B.M.W., Barlinn, K. and Siepmann, T., 2016. Randomized controlled trials–a matter of design. *Neuropsychiatric disease and treatment*, 12, p.1341.
- Stanley, K., 2007. Design of randomized controlled trials. Circulation, 115(9), pp.1164-1169.
- Stoltzfus, J.C., 2011. Logistic regression: a brief primer. Academic emergency medicine, 18(10), pp.1099-1104.
- Tyupa, S., 2011. A theoretical framework for back-translation as a quality assessment tool. *New Voices in Translation Studies*, 7(1), pp.35-46.
- Yen, M. and Lo, L.H., 2002. Examining test-retest reliability: an intra-class correlation approach. Nursing research, 51(1), pp.59-62.
- Yong, A. G. & Pearce, S., 2013. A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), pp. 79-94.
- Yusoff, M.S.B., 2019. ABC of content validation and content validity index calculation. *Resource*, 11(2), pp.49-54.
- Watson, R., 2015. Quantitative research. Nursing Standard, 29(31), p.44.
- Warren, M., Lininger, M.R., Chimera, N.J. and Smith, C.A., 2018. Utility of FMS to understand injury incidence in sports: current perspectives. Open access journal of sports medicine, 9, p.171.

ADVANCED QUANTITATIVE RESEARCH METHODS IN NURSING

LUCIJA GOSAK ET AL.

University of Maribor, Faculty of Health Sciences, Maribor, Slovenia lucija.gosak2@um.si

The publication "Analysis of quantitative research data in nursing research: A guide to SPSS" provides nursing students and nurses with the knowledge and skills to interpret the different statistical methods in their field, which can improve users' skills in collecting, analysing and interpreting results from clinical practice, thus contributing to improving the quality of health care. It provides detailed instructions on how to use IBM SPSS and perform statistical analyses that nurses need to be familiar with as they use and generate data in their daily work with patients. The main aim of patient care is to provide high quality, evidence-based care, so nurses have a duty to keep up to date with the latest research and evidence and apply it to their work. The knowledge gained in this book can also help nurses to better understand and interpret previously published results, and thus critically assess the validity and reliability of the results they will use in clinical practice. DOI https://doi.org/ 10.18690/um.fzv.2.2024

> **ISBN** 978-961-286-888-8

Keywords: uantitative analysis, statistics, IBM SPSS, reliability, validity, data analysis





Faculty of Health Sciences

