

# Statistical Product and Service Solution (SPSS) Essentials for Sports Data Analysis

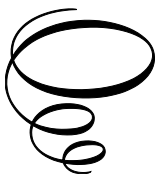


# Statistical Product and Service Solution (SPSS) Essentials for Sports Data Analysis

By

Rocsana Bucea-Manea-Țoniș

**Cambridge**  
**Scholars**  
Publishing



Statistical Product and Service Solution (SPSS)  
Essentials for Sports Data Analysis

By Rocsana Bucea-Manea-Țoniș

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Copyright © 2026 by Rocsana Bucea-Manea-Țoniș

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 978-1-0364-6850-7

ISBN (Ebook): 978-1-0364-6851-4

# CONTENTS

Foreword .....	ix
Recommendation Letter by Professor Gheorghe Orzan .....	xii
Chapter 1 .....	1
SPSS – Getting started	
Chapter 2 .....	19
Descriptive statistics	
2.1. Variables types.....	20
2.2. Scale types .....	24
Chapter 3 .....	40
Inferential statistics	
3.1. Ratio or Interval Scale.....	43
3.1.1. One Sample T test .....	43
3.1.2. Independent sample T test.....	45
3.1.3. Paired Sample T test.....	46
3.2. Nominal scale.....	48
3.2.1. Chi-Square Test / Fisher test .....	48
3.3. Ordinal scale .....	53
3.3.1. Sign Test.....	53
Chapter 4 .....	57
Analysis of Variance	
4.1. Interval or Ratio Scale – One-Way Anova.....	59
4.2. Ordinal/nominal scale - Kruskal-Wallis 1 Way Anova.....	63
4.3. Interval or Ratio Scale -Two-Way Anova.....	66
4.3.1. Two Way Anova with interaction without replication .....	72
4.3.2. Two Way Anova with replication .....	79
4.4. Ordinal scale or nonparametric variables - Wilcoxon.....	88
4.5. AnCova .....	91
4.6. MANova – General linear multivariate model.....	95
4.7. MANCova – General linear multivariate model .....	101

Chapter 5 .....	109
Analysis of the statistical association	
5.1. Nominal scale - Contingency coefficient .....	111
5.2. Ordinal scale .....	118
5.2.1. Spearman correlation coefficient.....	118
5.2.2. Kendall correlation coefficient .....	121
5.3. Interval/Ratio Scale – Pearson Correlation Coefficient .....	123
Chapter 6 .....	128
Regression	
6.1. Simple linear regression.....	131
6.1.1. Test Assumptions .....	134
6.1.2. Regression Statistics .....	136
6.1.3. Graphs Frequently Used in Regression .....	137
6.2. Multiple linear regression – stepping method .....	144
6.3. Partial regression of the smallest squares.....	151
6.4. Curve estimation .....	153
6.5. Logistic regression .....	156
Chapter 7 .....	159
Analysis of the statistical model reliability - Cronbach's Alpha	
7.1. Reliability and validity of the construct .....	164
7.2. Correlation between variables.....	165
7.3. Discriminatory validity .....	166
7.4. Model Fit.....	166
7.5. Multicollinearity analysis.....	166
Chapter 8 .....	173
EFA - Exploratory Factor Analysis	
8.1. Criteria for choosing the type of rotation .....	175
8.1.1. Orthogonal rotations.....	176
8.1.2. Oblique rotations .....	177
8.2. Interpretation of the result.....	177
8.3. Exemplification.....	178
Chapter 9 .....	191
Cluster Analysis	
9.1. What is clustering?.....	192
9.2. Types of grouping methods.....	193
9.2.1 Connectivity-based clustering (hierarchical grouping).....	194

9.2.2. Centroid grouping or partition .....	199
9.2.3. Density-based grouping (model-based methods) .....	200
9.2.4. Distribution-based grouping .....	202
9.2.5. Fuzzy grouping.....	203
9.2.6. Coercion-based (supervised grouping).....	203
9.3. Types of grouping algorithms .....	204
9.3.1 K-Means Clustering .....	205
9.3.2. Average algorithm for grouping displacements .....	209
9.3.3. Gaussian mixture model.....	210
9.3.4. DBSCAN.....	211
9.3.5. BIRCH algorithm .....	214
9.4. Applications of clustering .....	214
 Chapter 10 .....	 217
Time series	
10.1. Dynamic Time Series.....	220
10.2. Forecasting methods .....	222
10.2.1. Methods of extrapolation.....	223
10.2.2. Adjustment methods.....	223
10.2.3. Choosing the adjustment model .....	225
10.3 ARIMA .....	230
10.4. Seasonal decomposition.....	231
 Chapter 11 .....	 252
Neural Network Analytics	
11.1. Introduction to Neural Networks .....	252
11.1.1. Neural Network Architecture .....	253
11.2 Steps to Use Neural Networks in SPSS .....	254
11.3. Neural Network Models in SPSS.....	258
11.3.1. Multilayer Perceptron (MLP).....	258
11.3.2. Radial Basis Function (RBF).....	259
11.4. Output Interpretation in SPSS.....	260
11.5. Case Study: Predicting Recovery Outcome .....	272
 Chapter 12 .....	 288
Augmenting Sports Data Analysis: Integrating SPSS with ChatGPT	
12.1. Why AI in Sports Data Analysis? .....	288
12.2. Using ChatGPT to Build Research Questions.....	290
12.3. ChatGPT-Assisted SPSS Syntax Creation .....	294
12.4. Interpreting Output Using ChatGPT .....	294
12.5. Writing Research Results with AI Support.....	294

12.6. AI Ethics and Verification .....	297
12.7. Lab Activity .....	299

# FOREWORD

The link between the contents of statistics, occupational and professional standards, as well as the requirements of the labor market in the context of a doctoral school of sport and kinesiotherapy is essential for the training and development of the necessary skills in these fields. Here are some key points that underline this relationship:

## 1. The role of statistics in sports research and practice and physical therapy

Statistics play a central role in scientific research, especially in applied fields such as sports and physical therapy. Knowledge of statistics is necessary for:

- Analysis of experimental data: Evaluation of the effectiveness of training programs, studies on athletic performance, analysis of injuries, etc.
- Validation of recovery interventions: In physical therapy, the application of rigorous methods to evaluate the results of treatments is crucial.
- Development of scientific research: A PhD student must have the ability to collect, analyze and interpret data in order to contribute to scientific knowledge in the field of sport and health.

## 2. Occupational and professional standards

Within the sports and physical therapy professions, there are specific occupational and professional standards, which outline the requirements for the skills that a specialist must have. Statistics are directly linked to these standards by:

- Evaluation and monitoring of professional performance: Use of metrics and statistical indicators in the evaluation of professional activity (sports performance, effectiveness of treatments).
- Professional certifications: The requirements for obtaining certifications or accreditations in sports and physical therapy often include basic knowledge in data analysis and statistics.

### 3. Labour market requirements

The job market in sports and physical therapy places an increasing emphasis on skills based on data analysis, which are grounded in statistics. Market requirements include:

- Assessment and monitoring skills: Employers are looking for professionals who can assess the effectiveness of rehabilitation programmes or training plans using statistical methods.
- Innovation and applied research: To remain competitive, professionals must be able to use statistics to develop new and more effective solutions.
- Trends in big data and sports: The emergence of technology and the collection of large volumes of data from sports (performance, biomechanics, sports psychology) makes it necessary to use advanced statistics for the analysis of this data.

### 4. Integrating statistics into the doctoral school curricula

In doctoral schools of sport and physical therapy, statistics are commonly included in the training program to ensure the development of research skills. This includes:

- Methodology of scientific research: Doctoral students are trained to use statistical methods in the design of their studies, as well as in the interpretation of results.
- Use of statistical software: Students are often familiar with statistical tools such as SPSS, R, or other data analysis platforms, necessary in their research.
- Courses dedicated to data analysis: Within the doctoral programs, specialized courses are taught on topics such as multivariate analysis, inferential and descriptive statistical tests.

### 5. Specific requirements for sports and physiotherapy professionals

The labor market demands more and more specialization in the field of statistics and data analysis for sports and physical therapy professionals. Examples of such requirements include:

- Sports Performance Analysis: The ability to collect and interpret data on athletes' performance, injuries, and recovery.

- Personalization of rehabilitation plans: Using statistics to tailor treatments based on individualized patient data.
- Modern technologies and biometric data: The increasing use of real-time monitoring devices for athletes and patients (such as wearable devices) requires an advanced understanding of data analytics.

The statistics contents within the doctoral school of sport and kinesiotherapy are closely related to occupational and professional standards, but also to the requirements of the labor market. Statistical knowledge equips graduates with essential research skills and the ability to respond to practical challenges in sport and rehabilitation, giving them a competitive advantage in professional development and contributing to innovations in the sector.

# RECOMMENDATION LETTER

GHEORGHE ORZAN

It is with great pleasure and professional confidence that I recommend the publication of the book entitled:

**"SPSS Essentials for Sports Data Analysis: From Descriptive to Predictive Analytics"** *Authored by Dr. Rocsana Bucea-Manea-Țoniș*

This work stands out through its structured approach to statistical analysis tailored for sports science, kinesiotherapy, and related applied fields. Covering a wide range of techniques—from foundational descriptive statistics to advanced methods such as regression models, factor analysis, cluster analysis, and time series forecasting—the book integrates both theoretical depth and practical application using IBM SPSS.

What makes this book particularly valuable is its educational clarity and alignment with occupational and academic standards. It offers not only methodological guidance but also direct SPSS interface references and illustrations that make complex statistical concepts accessible, especially for doctoral students and professionals in sports performance and rehabilitation sciences.

Dr. Bucea-Manea-Țoniș has succeeded in creating a resource that bridges academic rigor with usability, suitable for both research and applied contexts. Moreover, the integration of AI tools, such as ChatGPT for syntax generation and results interpretation, reflects a modern, forward-thinking pedagogical vision.

I strongly support the dissemination and academic recognition of this volume and recommend it as a reference work for students, educators, and professionals who seek to enhance their data analysis competencies in sports-related disciplines.

Sincerely, **Prof. Gheorghe Orzan, PhD**  
Faculty of Marketing, Bucharest University of Economic Studies  
Senior Academic and Methodological Advisor

# CHAPTER 1

## SPSS - GETTING STARTED<sup>1</sup>

Motto: *"Numbers have an important story to tell. They rely on you to give them a clear and convincing voice."*  
Stephen Few (Data Visualization Expert).

Any processing can be done in extremely varied conditions. For example, the calculation of frequencies may or may not be accompanied by the display of the frequency table, by the graphical representation by histogram or by column chart, by the calculation of statistical indices such as quartiles, indices of the central tendency (mean, median, mode), measures of dispersion (variance, standard deviation, minimum and maximum value); the graphical representation can take into account absolute frequencies of values or percentages; the normal curve may or may not overlap the histogram. SPSS can achieve various processing variants.

SPSS provides extensive help resources to know how to perform a analysis. The *Help* option is permanently accessible and can explain the meaning of various terms in menus and dialog boxes.

In addition, an online tutorial program provides explanations and examples that help to quickly orient yourself among the numerous processes that can be carried out by this product.

SPSS can also be used by people who are not very initiated in statistics. For any notion, from a dialog box or even from an output list, you can get the display of an explanatory text if you indicate that name, on the screen, with the mouse and choose from the menu that the option *What's This?*

With SPSS you can even update and complete your knowledge of statistics, as the *Help* button allows access to a glossary of statistical terms.

---

<sup>1</sup> Footnotes.

Entering and modifying data is a simple operation, thanks to the existence of a spreadsheet-like editor. A table is displayed on the screen, whose lines correspond *to the cases* (subjects answering a questionnaire or objects observed) and in whose columns *there are the variables* (answers given by the subjects or the results of measurements or observations - *see figure 1.1*). The user can "navigate" through this table, as desired, inspecting the existing values, changing some data, adding or deleting cases or variables. SPSS automatically adapts the dimensions of the table so that no input value is lost. There are no limitations on the number of variables or cases.

### **Automatic transformation of input data**

Before performing statistical processing, the data can be automatically modified by SPSS, based on algorithms indicated by the user for recoding the values or by applying mathematical functions. For example, in a file where the cases are represented by various goods, and their characteristic variables, all prices can be changed by adding VAT, or all goods produced before 2000 can be given the same value of the date of manufacture, meaning "before 2000".

### **Selection of processed cases**

The user can choose the cases to be taken into account when performing the processing, formulating conditions on the values of one or more variables. For example, in the study of the relationship between the level of schooling and the preference for various musical genres, only people residing in urban areas can be included from the entire data file.

The user can decide how to deal with cases where the value of a variable is not known or is not of interest to the research by the SPSS. They may or may not be included in the calculations.

Various graphs allow you to quickly detect outliers. These are isolated values, very different from the others in the file, which could be due to errors in data entry.

### **Control over the display of information**

The user has full control over all processed variables. He decides how to display the values in the output lists (on what length, with how many decimal places) and what text to write instead of variable names (if they

are not suggestive enough) or instead of variable values (if codes have been entered in the data file). For example, if for the variable corresponding to the performance in research, with the name "Perform\_senz", the values -2, -1, 0, 1, 2 have been entered in the file, the output lists may be written "Totally disagree" instead of -2, "Disagree" instead of -1, "Neutral" instead of 0, "Agree" instead of 1 and "Totally agree" instead of 2 (Fig.1.1.).

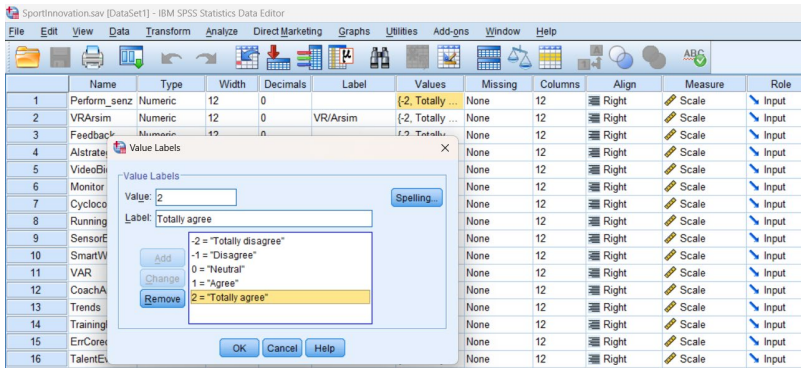


Fig. 1.1. Display of cases and variables

Choosing how to process the data is particularly simple, thanks to the menus and dialog boxes that can be reached by operating the mouse. For users who do not like to use the mouse, there is the possibility of indicating the desired processing by writing commands.

If a processing group is carried out periodically (for example, if the daily situation of sales by stores and products is of interest), there is no need to repeat the sequence of menu searches and choice of options every time. SPSS can be asked to record all these operations in a command file, executed by the user only once. The file will be run whenever necessary. If necessary, the file can be modified by adding or deleting commands.

The results of statistical processing can be visualized through tables of various formats (16 table styles are possible) and through many types of graphical representations: histograms, column charts - isolated or grouped, circular structure charts, statistical clouds in which the points corresponding to different groups of cases are colored differently, charts that indicate the average at the same time, the extreme values and the distribution of the values of a variable for different values of this variable

(e.g. the graphical representation of the age of the people, depending on the locality of residence).

## **Results are subject to change**

The list of results, tables and graphs made by SPSS can be included in the reports, as presented on the screen or modified.

The user can intervene in them by:

- Text editing;
- Change font characteristics (type, style, color, size);
- Modifying drawings by moving the axes, rotating them or even changing the type of graphic;
- Hiding variables from tables;
- The reorganization of the information in the tables (for example, a frequency table containing the answers to a questionnaire on lines, and the place of residence and, with in each locality, the gender of the customers, can be transformed into a table with only two columns, corresponding to the sex and with groups of lines, one group for each locality).

All these operations are easy to execute, thanks to the existence of three editors: text, tables and graphs. For example, changing the place of a column in a table is done by "dragging" the mouse to the icon corresponding to it.

## **Launch into SPSS Execution**

The SPSS program opens by following the path: START/PROGRAMS/SPSS 29 FOR WINDOWS. When opened, a window with several options appears (fig. 1.2):

- New Files – opening a page for manual data entry or launching a program with the formulation of a new query
- Recent files (program launch by opening a data source file with content previously filled in with the extension ".sav").
- The SPSS program opens an interments window, which allows the export of SPSS graphs and tables, presents the installed SPSS modules, provides tutorials for using the various modules (Fig.1.2.).



Fig. 1.2. Launch of the running program

## SPSS Modules

SPSS is made in modular form. Consequently, each user has the opportunity to purchase only those components that are necessary for him. The main modules sold are presented below.

- The Analyze module performs data and file management, data transformation, various graphical representations (histograms, structure diagrams, statistical clouds, etc.).  
Frequently used statistical processing are:
  - Calculation of frequencies, central tendency indices and variability indicators;
  - Calculation of measures of association and testing of probable independence for data included in contingency tables;
  - Comparison of sample means and variances;
  - Unifactorial analysis of variance;

- Calculation of the Pearson, Kendall and Spearman correlation coefficients:
  - Linear regression analysis;
  - Nonparametric tests.
- The Regression and Advanced Statistics modules include the performance of complicated statistical processing, necessary when the data do not meet the conditions of application of the usual (simpler) statistical processing and allow the execution of procedures that serve to investigate the relationships between variables, through the following methods:
- Discriminant analysis;
  - Factor analysis;
  - Cluster analysis;
  - Multidimensional scaling;
  - Weighted regression;
  - Reliability analysis (e.g., Cronbach's alpha)
  - Logistic regression analysis;
  - Various extensions of the ANOVA unifactorial variant analysis;
  - MANOVA multivariate variant analysis;
  - Loglinear analysis;
  - Nonlinear regression analysis;
  - Probit and logit analysis;
  - Survival analysis (including Kaplan–Meier estimation).
  - General Linear Model (GLM).
- The Tables module is used to condense data into one, two, or three-dimensional tables. Each dimension is defined by a variable or by a group of variables (for example, the lines can represent the profession and gender of the subjects, and the columns, the possible answers to an item in a questionnaire). The tables may contain, along with the values of the variables, frequencies and values of some statistical indicators (mean, standard deviation, etc.).
- The Exact Tests module determines the significance level ("p") for:
- Nonparametric tests applied to one sample, to two independent or paired samples and to k dependent or independent samples;
  - Tests applied to contingency tables 2\*2 and r\*c;
  - Significance tests for Pearson and Spearman correlation coefficients;

- Tests related to the relationships between variables measured on nominal or ordinal scale.

This module calculates exact values for the significance levels of statistics, unlike the basic module, which only calculates asymptotic values. Exact tests are particularly useful for small samples or sparse contingency tables.

- The CHAID module (Chi-squared Automatic Interaction Detector) mode applies segmentation algorithms to divide a population into disjunct groups, which differ from each other in terms of a specified criterion. The groups formed at each step of the algorithms are visualized in the form of a dendrogram.
  - The Categories mode is used to determine the influence exerted by the characteristics of products or services on consumer preference. It also allows the visualization of the analyzed objects, through points, in order to identify the similarity or difference between them.
  - The Trends mode is used for the analysis and graphical representation of time series. It estimates the coefficients of the model according to a time series, using the following techniques:
    - Regression methods;
      - Box-Jenkins analysis (ARIMA), for one-dimensional seasonal and non-seasonal models;
      - Seasonal decomposition procedure, for determining additive and multiplicative seasonal factors, from periodic time series;
      - Analysis of frequency components.
- Further details can be found on the website <http://www.spss.ro/>.

## SPSS Menu\*

**File** - includes the operations specific to any Microsoft program, namely opening an existing file, creating and saving a new file, importing files with different formats (Excel, Fox pro, Access, SQL), viewing the information in the file opened before listing and listing it.

**Edit** - similar to the File menu and this menu follows the structure of Microsoft programs, comprising the editing commands in the file: copy,

---

\*Gheorghe Orzan - "Marketing Information Systems", Uranus Publishing House, 2001

paste, move the general data of the program, the operation to find a certain value within a variable and the general options of the program (related to graphs, pivot table, currency, scripts, labels, data type, etc.)

**View** - for viewing certain elements: status bar, toolbar, font and text features, viewing labels and/or variable content, and page grid.

	Perform_senz	VRArsm	Feedback	Alstrateg	VideoBio	Monitor	Cyclocomputer	RunningPod	SensorBall	SmartWatch	VAR	CoachApps
1	2	2	2	2	2	2	1	1	1	1	1	1
2	1	1	1	0	1	0	2	2	2	2	2	2
3	2	1	2	0	2	0	2	1	0	1	1	2
4	1	0	2	1	1	1	2	1	1	2	2	0

Fig. 1.3. SPSS Menu

**Data** - for modifying data within the database. You can define the characteristics of different variables, including date type, you can make templates, insert new variables, new cases. Other operations refer to the operation of inversion (transposition) of the matrix that forms the database, sorting the cases according to several criteria, creating variables: mean, sum, minimum or maximum value, standard deviation, etc. You can merge (concatenate) several files. Two more important operations refer to the selection of cases according to certain criteria and the recovery of the sample.

**Transform** - in which a new variable can be generated as a result of operations between variables, or functions applied to these variables; time series can be generated or missing values can be substituted.

**Analyse** - which contains all the statistical procedures used by the SPSS. These statistical procedures will be exemplified in the following chapters (Fig. 1.4).

**Direct Marketing** - which contains all the procedures for creating customer profiles and carrying out marketing campaigns (Fig. 1.5.).

**Graphs** - for generating graphs in two- and three-dimensional space using a wide range of graphic types and patterns.

**Utilities** - to get data related to variables, file, to run macros or to customize the SPSS menu.

**Add-On** – contains a wide variety of applications, services, and extensions for programming in SPSS. Among these we mention AMOS, which allows Path analysis. Different licenses are paid for these extensions.

**Window** - to switch from one window to another, SPSS using several windows that will be shown below.

**Help** - contains on-line information about all modules and operations used by SPSS, links to the SPSS website and to the syntax manual.

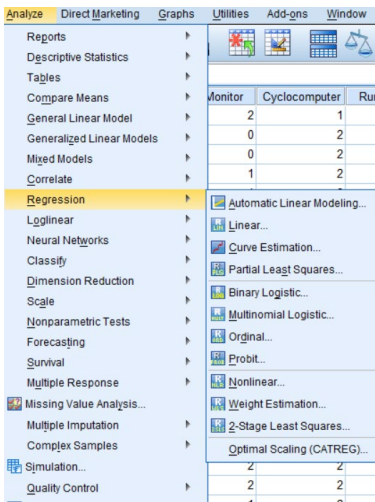


Fig. 1.4. Analysis menu

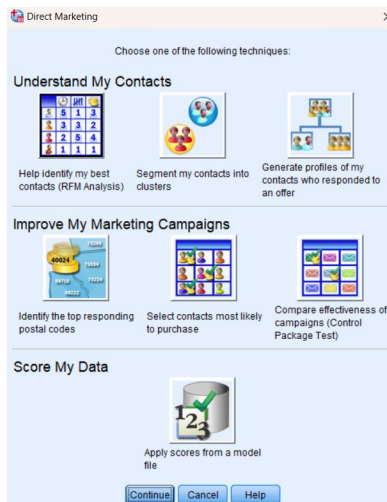


Fig. 1.5. Direct Marketing Menu

## SPSS Windows

- **Data Editor** – is the first window that welcomes the user when starting work sessions and allows viewing information from existing files or creating new files. This window has the appearance of an Access table or an Excel page, in which each column represents a variable (detached from the questionnaire or from the characteristics of a reality to be studied), and the records are entered in rows (different values of the variable, given by interviewees or specialists who can form a focus group; they are observations);

- **Output Viewer** – following the analysis procedures, SPSS automatically displays the results in this window. Here you can view: the name of the procedure performed, the levels of meaning, graphic or tabular representations on which different formatting can be performed (from choosing the table format, the font type, changing the title, to the different arrangement of the information by pivoting), (Fig. 1.6);

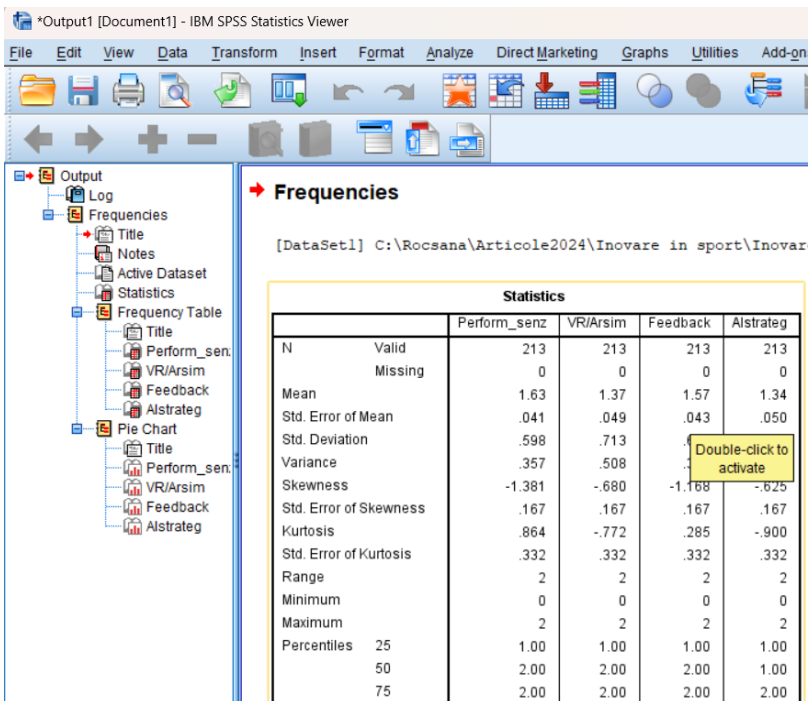


Fig. 1.6. Output Viewer window

- **Draft Viewer** - used to present results in text form, instead of interactive pivot tables;
- **Pivot Table Editor** - the results presented in the form of pivot tables can be modified in many ways with the pivot table editor. You can edit text, transpose rows into columns and vice versa, add color, create multidimensional tables, and hide or expose results.
- **Chart Editor** – in this window you can make changes regarding the graph (color, type, font and content of the variable labels or the

title of the graph, change the horizontal and vertical axes, its rotation angle), but also changes regarding the variables and the type of analysis applied to them that can be changed from the **Analyze** menu (Fig. 1.7);

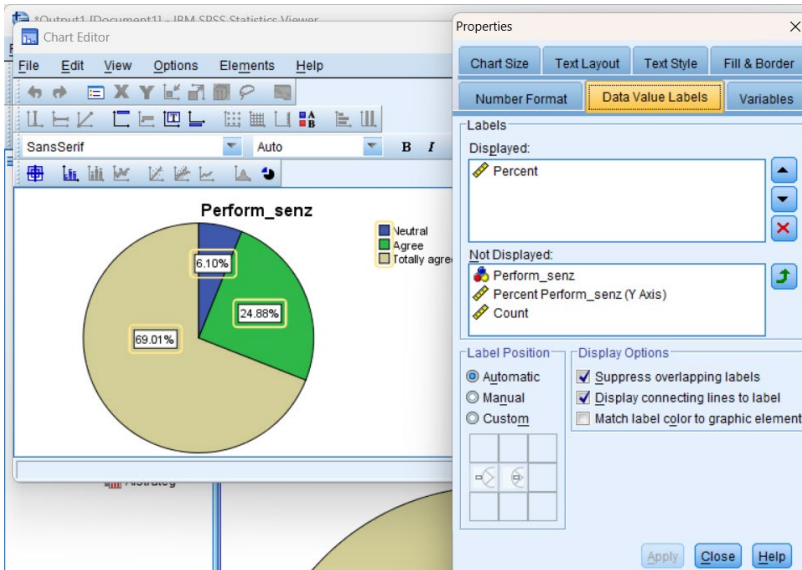


Fig. 1.7. Chart Editor window

- **Text Output Editor** – window for modifying the results in the form of text contained in pivot tables. You can edit the results and change the font characteristics;
- **Syntax Editor** - all menu operations can be written as a command line and edited, benefiting from some features and features of the program that are not presented in the menus. These syntaxes can be saved for later use;
- **Script Editor** - automates commands through scripting and OLE features, allowing custom functions to be created in SPSS.

## Encoding and data entry

An example of marketing analysis can be made based on an "SPSS sample" that contains as variables the answers to the questions of a

questionnaire applied in an opinion poll in the United States. Among the questions are:

1. What region do you come from?-the variable "region" is a qualitative variable, measured on a nominal scale.
2. How do you characterize your level of happiness? - "happiness" is a qualitative variable measured on an ordinal scale, if the intervals are not equal (low, medium, high) or an interval scale if we use equal intervals such as Likert scale (1= very low, ..... 5=very high). Although Likert items are ordinal, many researchers treat composite Likert scales as approximately interval for parametric testing.
3. What is your gender? -"gender" is a qualitative variable, measured on a nominal scale.
4. What is the number of brothers/sisters? - "brothers" is a discrete quantitative variable (it cannot take values with commas - no one can have 1.5 brothers).
5. How old are you? - "age" is a quantitative variable measured on an ordinal scale if grouped into categories (e.g., 18–25, 26–35), or on a ratio scale if measured in exact years (e.g. your age is ...27... years old).
6. What is your occupation? - "occupation" is a qualitative variable measured on a nominal scale.

### **Variables – columns**

To insert a variable in an existing data source, right-click on the name of the variable before which we want to insert the new variable and select "Insert variable", for which we will choose a name (Fig. 1.8). This can be done in the Data View page, but also in the Variable View page

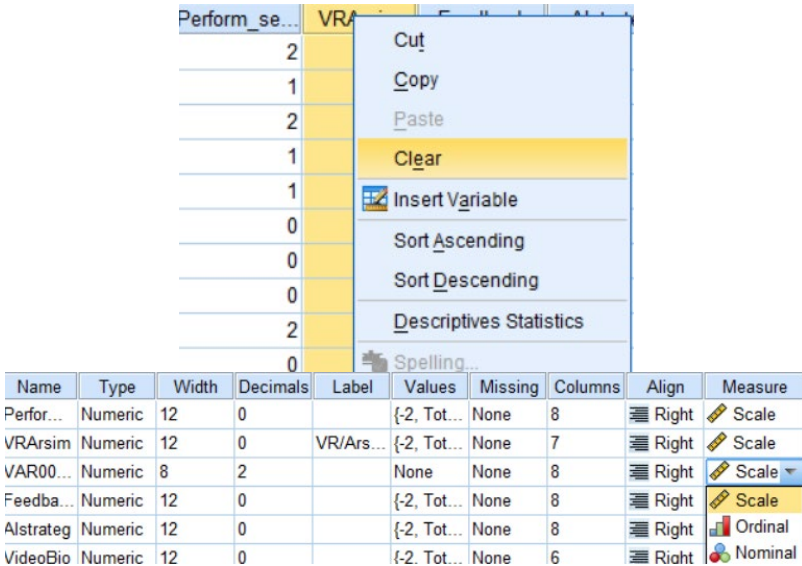


Fig.1.8. Inserting a new variable

Example: choose the variable "VRARsim" that corresponds to the question: "To what extent do you think VR/AR simulations enable innovation in sports?". The new variable will be inserted before it. For the new variable, all the characteristics are completed, the most important being the measurement scale (fig.1.8). Also from the secondary menu, by right-clicking on the name of a variable you can delete, copy or move this variable.

### Cases – rows

To enter a new case, i.e. the answers to all the questions of the questionnaire, given by a new person, the user positions the cursor and clicks on the number of the record in which he wants to enter the new case (the other cases will be moved with a record below). Activates **Insert Case** and fill in the person's answers in the Spreadsheet. With a right click on a case number, it can be deleted.

## Data menu

From the menu **Date** The cases can be rearranged ascending or descending, according to a certain criterion represented by a variable, by choosing the option "**Sort Cases**", the variables can be transposed, by choosing the option "**Transpose**", duplicates or different cases can be identified by choosing the "**Identify duplicate/ unusual cases**", you can choose calendar date formats, by choosing the option "**Define Dates**", you can merge or split data from different spreadsheets by choosing the "**Merge Files/Split Files**", a new spreadsheet can be created after applying an aggregate function for a certain variable by choosing the "**Aggregate**", you can filter the cases by a certain condition, filter, interval by choosing the option "**Select Cases**" (etc.).

## Recoding variables

If the number of cases is very different in each category, this problem can be corrected by regrouping the new categories. Balancing group sizes improves robustness of ANOVA, but does not correct violations of normality or homogeneity. Transformation or nonparametric tests may be more appropriate. One of the easiest ways to try to correct these inadequacies of the model to the assumptions of the method is the following:

- The categorical variable may be recoded so that group sizes are more balanced (this may improve robustness, but does not guarantee homogeneity of variance);
- The variable Y is normalized.

To recode the variable, open the menu **Transform – Recode into different variable**; choose the variable "salary"; rename it "salary" (fig. 1.9) – the new name will be "VARn" and press the **Change button**; press the **Old and New Values** button – in the Old values the range of values extracted from the old variable that will form a new category is passed to the range criterion, and in the New values quadrant a new value is given for the category from the newly created variable – press the **Add button**; resume the procedure for forming the new categories, press the **Continue** button and then the **OK button**. Salary data are typically right-skewed. Log transformation compresses high values, improving normality and stabilizing variance. When interpreting results, coefficients must be interpreted in the transformed (logarithmic) metric.

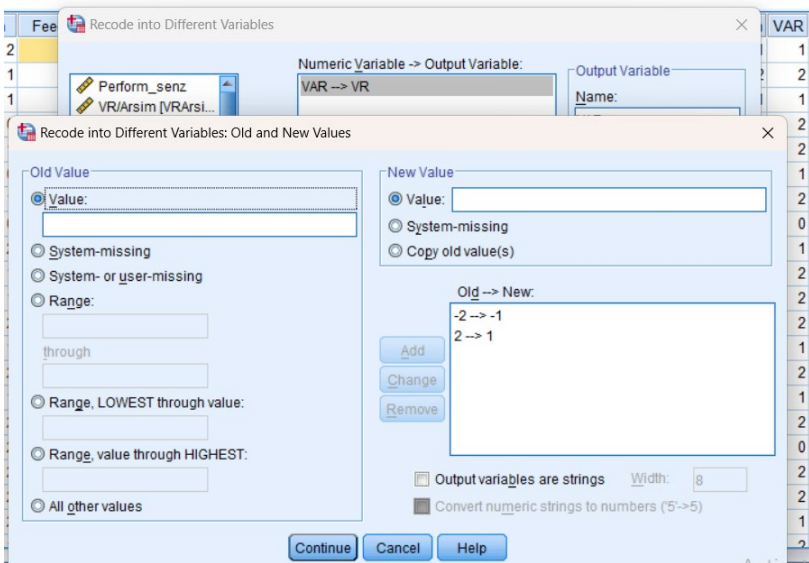


Fig. 1.9. Renaming and recoding the variable "VAR" in the new categories

If the variables do not have a normal distribution (a histogram of the distribution of the salary variable is constructed to observe the distribution) the numerical variable will be established. Depending on the shape of the distribution of the analyzed variable, a series of transformations are applied. In our case, the salary is logarithm. We construct a new variable that we call "Target Variable" *salnou*, and for "Numeric Expression" we specify that it is the natural logarithm of the salary variable (Fig. 1.10).

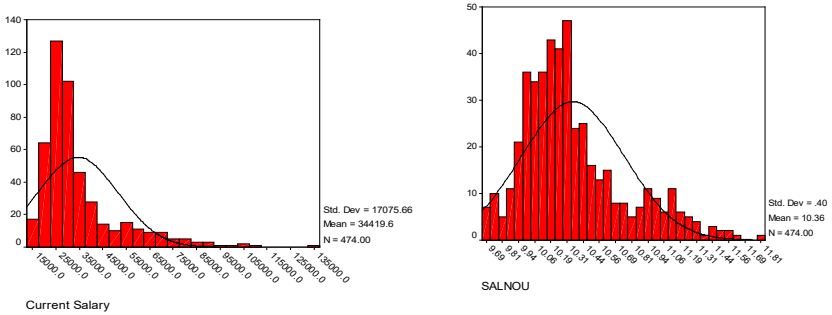


Fig. 1.10. a) Current salary and b) New salary

To normalize the variable, **open the Transform - Compute menu** (Fig. 1.11); give a name to the new variable in the Target Variable box – add the numeric expression "**ln(salary)**"; press the OK button.

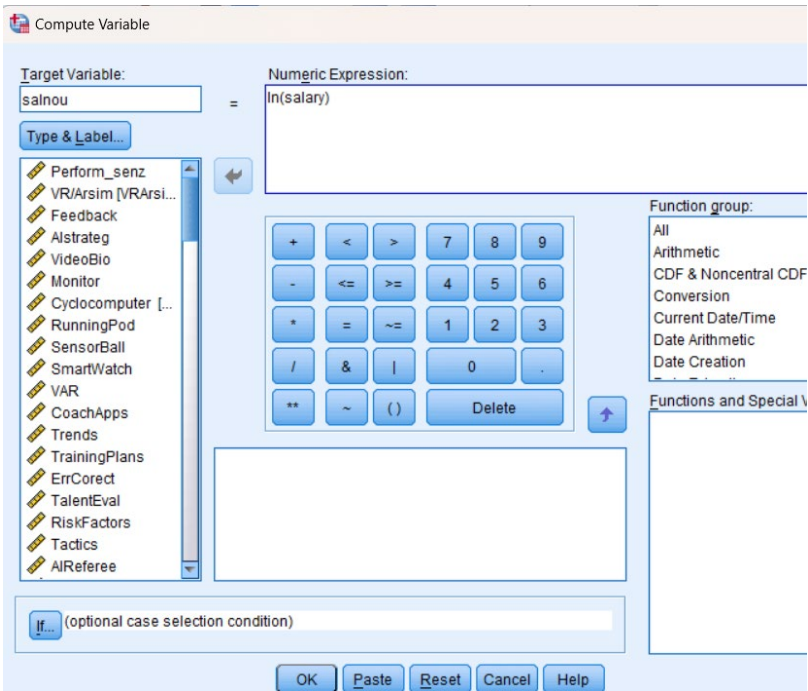


Fig. 1.11. Logarithm of the variable

The histogram shows that variable Y, logarithmic salary, is distributed approximately normally in the categories of variable X, level of education. (educs-recoded). We also apply the Levene test.

### Procedure:

- 1) Open the **Analyze- Compare mean - One Way Anova menu**;
- 2) choose the 2 variables - press the Options button;
- 3) check **Homogeneity of variance** - press the **Continue** button and then the **OK button**. Since  $p < .05$ , the assumption of homogeneity of variance is violated. One may therefore consider using Welch's ANOVA or applying a transformation. Alternatively, robust tests (e.g., Brown–Forsythe) may be used.

Test of Homogeneity of Variances - SALNOU

Levene Statistic	df1	df2	Sig.
15.764	2	451	.000

### Transformation of variables according to certain conditions

The variable aims to increase the salaries of young people (born after 1970) who worked for more than four years in the company. For this, the **Transform – Compute menu opens**; press the **If** button - check "**Include if cases satisfy condition**" - put the condition "jobtime >96 & bdate > 01/01/1970" - press the **Continue button**; give a name to the new variable in the **Target Variable box** – add the numeric expression "salary+50" and press the OK button. A new column is inserted in the SPSS Data Editor containing the modified salaries for the aforementioned persons.

### Selecting cases

For conditional selection, the Data - Select Cases menu opens; press the If button, put the condition "educ>15 & gender=2 & jobtime >96"; press the Continue button – press the OK button. Only those people who have more than 15 years of study, the female gender and have been working in the company for more than four years, are selected.

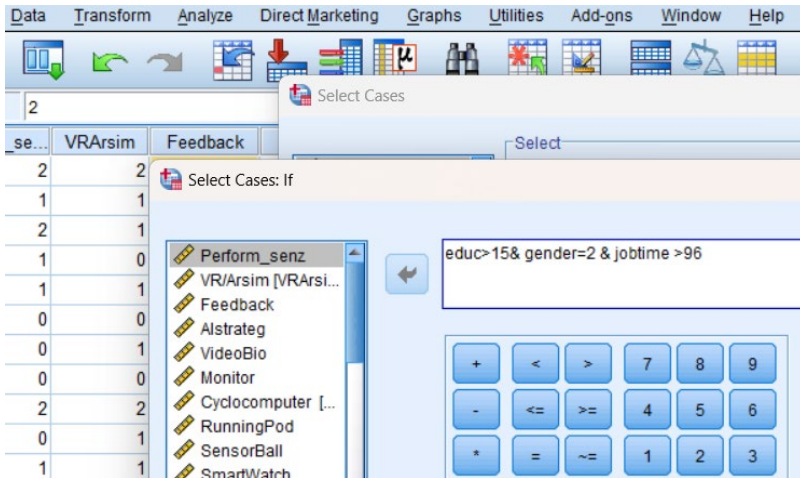


Fig. 1.12. Selecting variables

## Key Takeaways for Chapter 1: SPSS - getting started

- SPSS is a powerful tool for statistical analysis that simplifies data management, visualization, and inferential statistics.
- No advanced expertise in statistics is required for basic procedures - SPSS provides intuitive tools like menus, dialog boxes, and automated analysis to guide users.
- SPSS supports a wide range of statistical methods - from basic descriptive statistics to complex regression models and predictive analytics.
- Customizable output and reports allow researchers and analysts to present findings clearly and effectively.

## References

1. Few, S. (n.d.). *[Quote]*. Retrieved from <https://www.perceptualedge.com/>
2. IBM Corp. (2022). *IBM SPSS Statistics for Windows (Version 29.0)* [Computer software]. IBM Corp.
3. Orzan, G. (2001). *Sisteme informatice de marketing* [Marketing Information Systems]. Bucharest: Editura Uranus.
4. Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). London: Sage Publications.