PREDICTIVE ANALYTICS LAB

LAB MANUAL

B.TECH III YEAR – VI SEM(R23)

DEPARTMENT OF
COMPUTER SCIENCE AND ENGINEERING
(DATA SCIENCE)

SRINIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES

(Autonomous Approved by AICTE, New Delhi. Affiliated to JNTUA, Ananthapuramu)

SYLLABUS

COURSE EDUCATIONAL OBJECTIVES:

To develop the fundamental understanding and application of Mathematics and Statistics in business organizations

LIST OF EXPERIMENTS:

1. Introduction to SPSS, Sorting File, Split File, Compute File, Recode File and Select Cases
2. Chi-Square Test (Parametric and Non- Parametric)
3. Exploratory Factor Analysis
4. Cluster Analysis
5. Logistic Regression
6. Discriminant Analysis
7. Confirmatory Factor Analysis
8. Conjoint Analysis
9. Time Series
10. MANOVA
11. Decision Tree Analysis

TEXT BOOKS:

1. Andy Field, Discovering Statistics Using IBM SPSS Statistics, SAGE Publications.
2. George A. Morgan et al., IBM SPSS for Introductory Statistics: Use and Interpretation, Routledge.
3. Joseph F. Hair et al., Multivariate Data Analysis, Pearson.
4. Richard A. Johnson, Dean W. Wichern, Applied Multivariate Statistical Analysis, Pearson.

REFERENCE BOOKS:

1. Efraim Turban et al., Business Intelligence: A Managerial Approach, Pearson.
2. James R. Evans, Business Analytics: Methods, Models, and Decisions, Pearson.
3. Nitin R. Patel, D. P. Prajapati, SPSS for Data Analysis, McGraw-Hill.
4. R. A. Johnson, D. W. Wichern, Applied Multivariate Statistical Analysis, Pearson.
5. Joseph F. Hair et al., Multivariate Data Analysis, Pearson.

REFERENCE WEBSITE:

1. https://www.ibm.com/products/spss-statistics
2. https://statistics.laerd.com/
3. https://www.khanacademy.org/math/statistics-probability
4. https://www.coursera.org

EXP1: Introduction to SPSS, Sorting File, Split File, Compute File, Recode File and Select Cases

Aim: To study IBM SPSS Statistics software and to perform fundamental data management operations such as

- Understanding the SPSS interface
- Sorting cases in a dataset
- Splitting the file for separate group analysis
- Computing a new variable from existing variables
- Recoding values of a variable into a new variable
- Selecting specific cases for analysis

Introduction to SPSS

IBM SPSS Statistics (commonly referred to as SPSS) is a powerful and widely used statistical software suite developed for data management, advanced analytics, multivariate analysis, business intelligence, and predictive modeling. Originally standing for "Statistical Package for the Social Sciences," it was first launched in 1968. IBM acquired the software in 2009, and current versions (as of 2025, including versions around 29-31 and ongoing developments) are officially branded as IBM SPSS Statistics.

SPSS is designed to help researchers, analysts, organizations, and students extract actionable insights from data through:

- Robust statistical testing (e.g., descriptive statistics, t-tests, ANOVA, regression).
- Predictive analytics and machine learning algorithms.
- Data preparation, visualization (charts, graphs), and reporting.
- Integration with open-source tools like Python and R for extensibility.
- Handling complex data scenarios, including missing values, complex samples, and big data integration.

It features a user-friendly graphical interface with point-and-click menus, making it accessible for beginners while offering advanced options via syntax (command-language scripting) for experienced users. SPSS supports importing data from various formats (e.g., Excel, CSV, databases) and is used across fields like social sciences, healthcare, marketing, education, and government research.

Main Windows in SPSS

When you launch SPSS, the **Data Editor** window opens by default:

- **Data View**: Spreadsheet-like grid where rows represent **cases** (e.g., individual respondents or observations) and columns represent **variables**. This is where you enter, edit, and view actual data values.
- **Variable View**: Displays metadata (information about variables). Each row represents a variable, with columns for properties such as:

    - **Name**: Short unique variable name (no spaces, starts with letter).
    - **Type**: Numeric, String, Date, etc.
    - **Width/Decimals**: Display format.
    - **Label**: Descriptive variable label (can include spaces).
    - **Values**: Value labels for categorical data (e.g., 1 = "Male", 2 = "Female").
    - **Missing**: Codes for missing values.
    - **Columns/Align**: Display settings in Data View.
    - **Measure**: Level of measurement (Nominal, Ordinal, Scale).
    - **Role**: Input, Target, etc. (for certain procedures).

Switch between these views using tabs at the bottom-left of the Data Editor.

Other key windows:

- **Output Viewer**: Displays results of analyses (tables, charts, logs). Saved with .spv extension.
- **Syntax Editor**: For writing and running SPSS command syntax (advanced automation).

Organization of Options (Menu Bar)

The top menu bar provides access to all SPSS functions. Menus are logically grouped for workflow: from file handling and data manipulation to analysis and output.

Key menus and their main options:

- **File**: New/Open/Save data files; Import/Export data; Print; Recent files; Exit.
- **Edit**: Undo/Redo; Cut/Copy/Paste; Find; Options (preferences, e.g., output settings).
- **View**: Toggle toolbars, status bar, grid lines, value labels (in Data View); Menus (customize).
- **Data**: Define variable properties; Sort cases; Merge files; Aggregate; Split File; Select Cases; Weight Cases.

- **Transform**: Compute new variables; Recode variables; Categorize; Rank cases; Automatic Recode.
- **Analyze**: Core statistical procedures.
    - Descriptive Statistics (Frequencies, Descriptives, Explore).
    - Compare Means (t-tests, ANOVA).
    - General Linear Model; Correlation; Regression.
    - Nonparametric Tests; Scale; Multivariate.
    - Tables (Custom Tables); Forecasting; Survival; etc.
- **Graphs**: Chart Builder; Legacy Dialogs (bar charts, scatterplots, histograms); Graphboard Template Chooser.
- **Utilities**: Variables info; Data File Comments; Display Dictionary.
- **Add-ons**: Integration with extensions (e.g., Python/R); Manage custom dialogs.
- **Window**: Switch between open windows (Data Editor, Output, Syntax).
- **Help**: Tutorials; Documentation; Case studies; About SPSS (version info).

Procedure (Basic Startup):

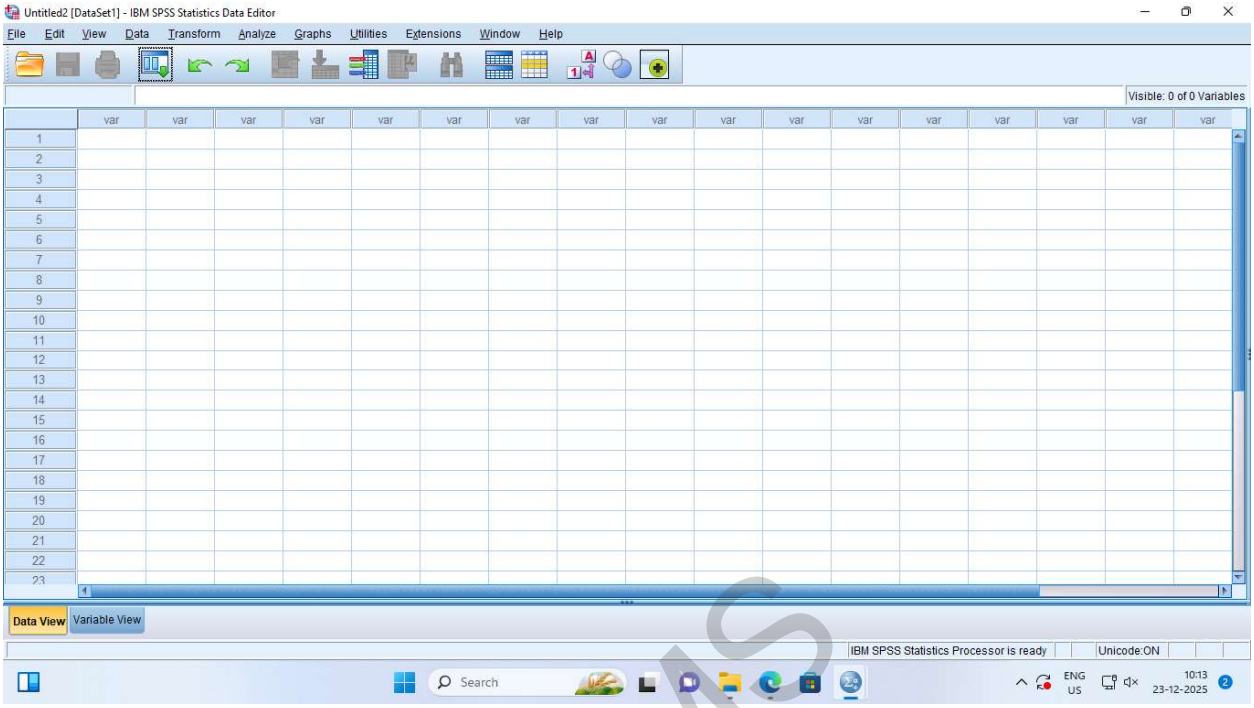
1. Launch IBM SPSS Statistics.
2. Explore the empty Data Editor (switch between Data View and Variable View).
3. Go to **File > Open > Data** to load a sample dataset (e.g., demo.sav if available).
4. Navigate the menu bar to familiarize with options.
5. Run a simple analysis: **Analyze > Descriptive Statistics > Frequencies** to see output in the Viewer.

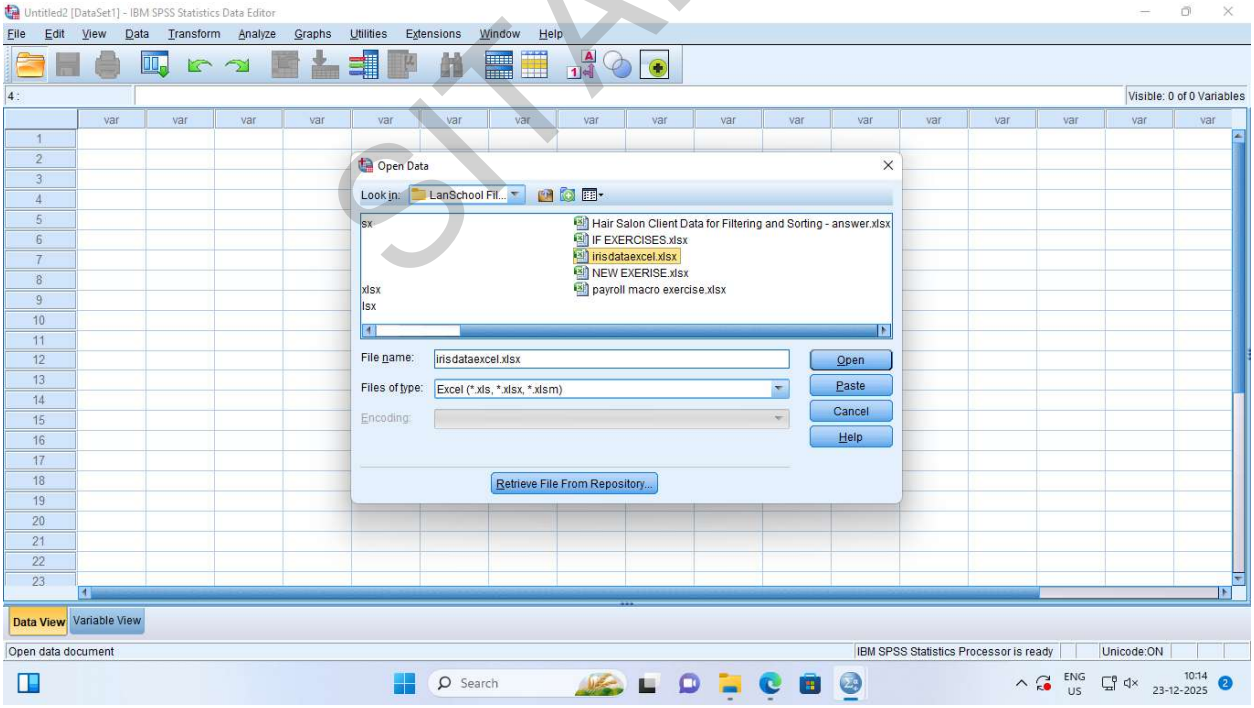
1.A) Sorting File (Sort Cases)

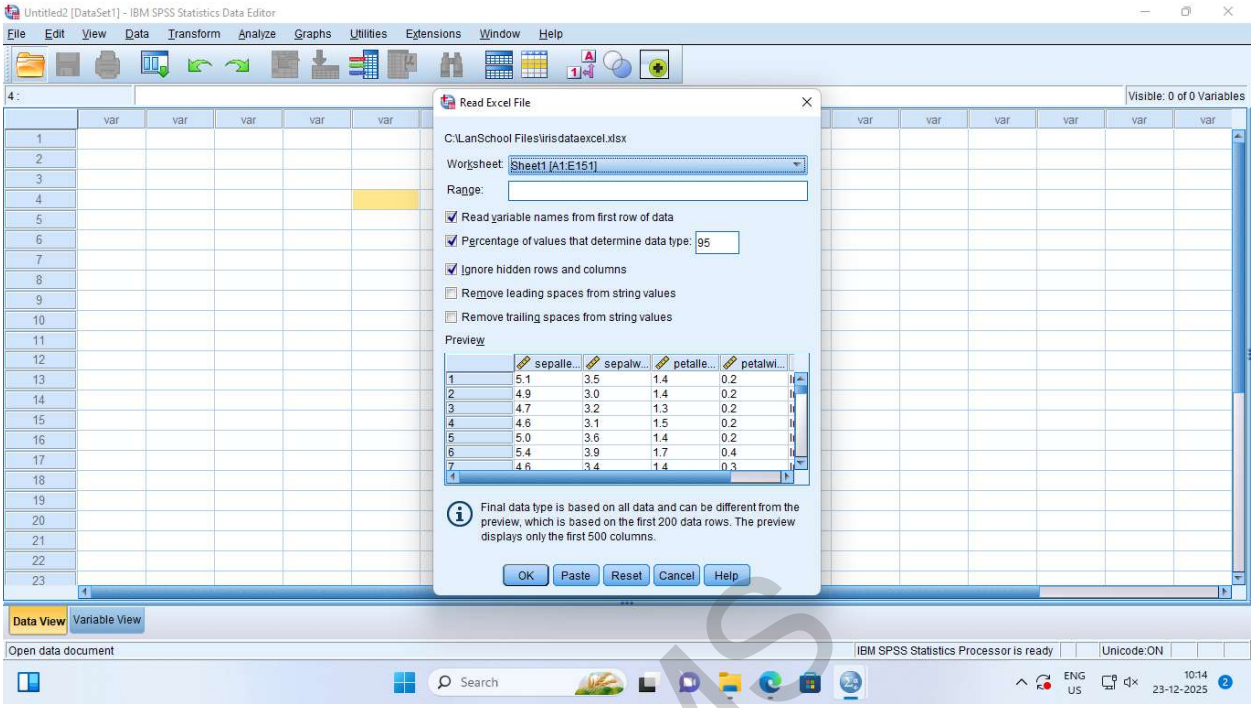
Aim: To rearrange the cases in the dataset based on the values of one or more variables in ascending or descending order.

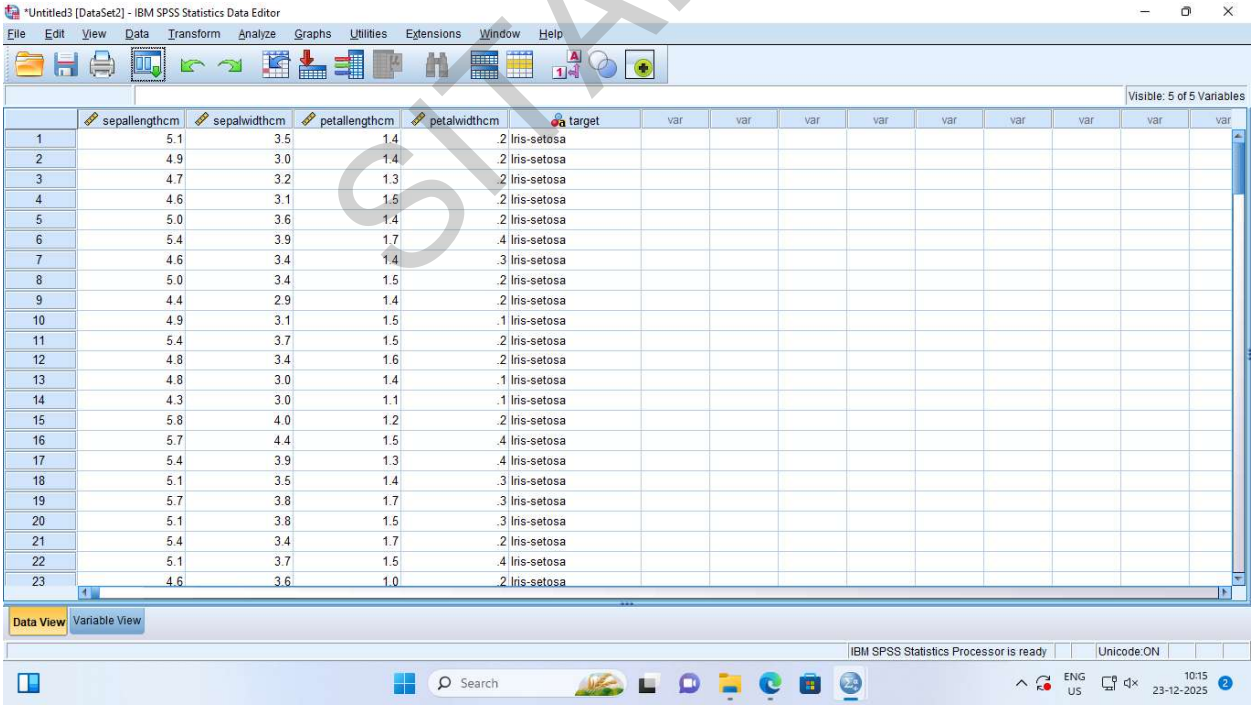
Procedure:

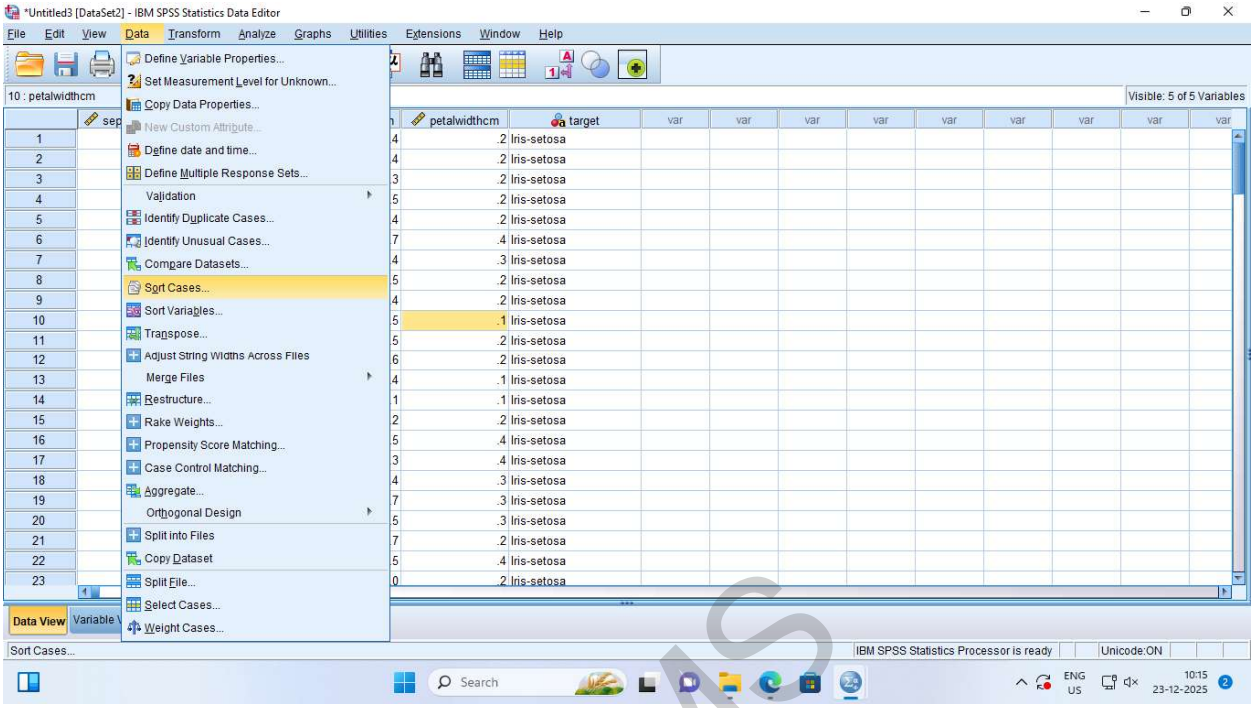
1. Open your dataset in the Data Editor.
2. Go to **Data > Sort Cases**.
3. In the Sort Cases dialog box, select the variable(s) you want to sort by from the left list and move them to the "Sort By" box using the arrow button.
4. For multiple variables, the primary sort variable is the first in the list, followed by secondary sorts.
5. Choose **Ascending** (default, low to high) or **Descending** (high to low) for each variable by selecting it and clicking the option.
6. Click **OK**.
7. The dataset in Data View will now be reordered. Note: This permanently changes the row order unless you undo or re-sort.
8. To return to original order, create an ID variable first (using Compute) before sorting.

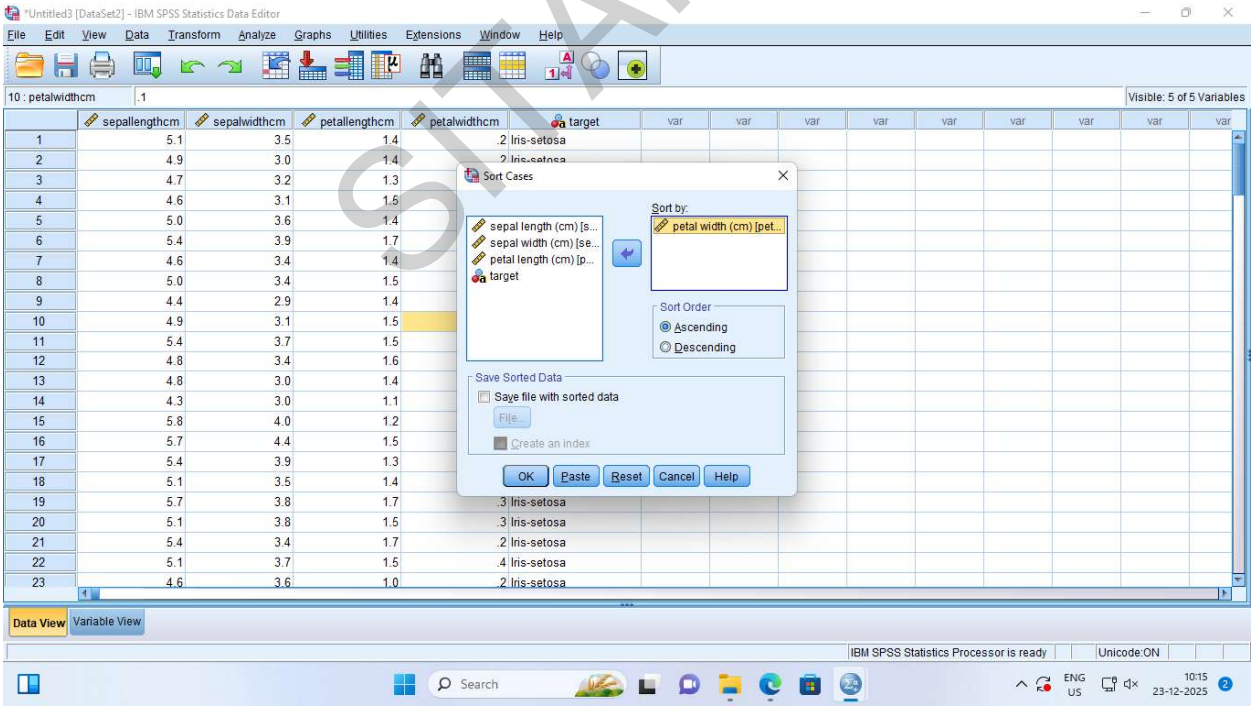

Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 0 of 0 Variables
var
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:13
23-12-2025

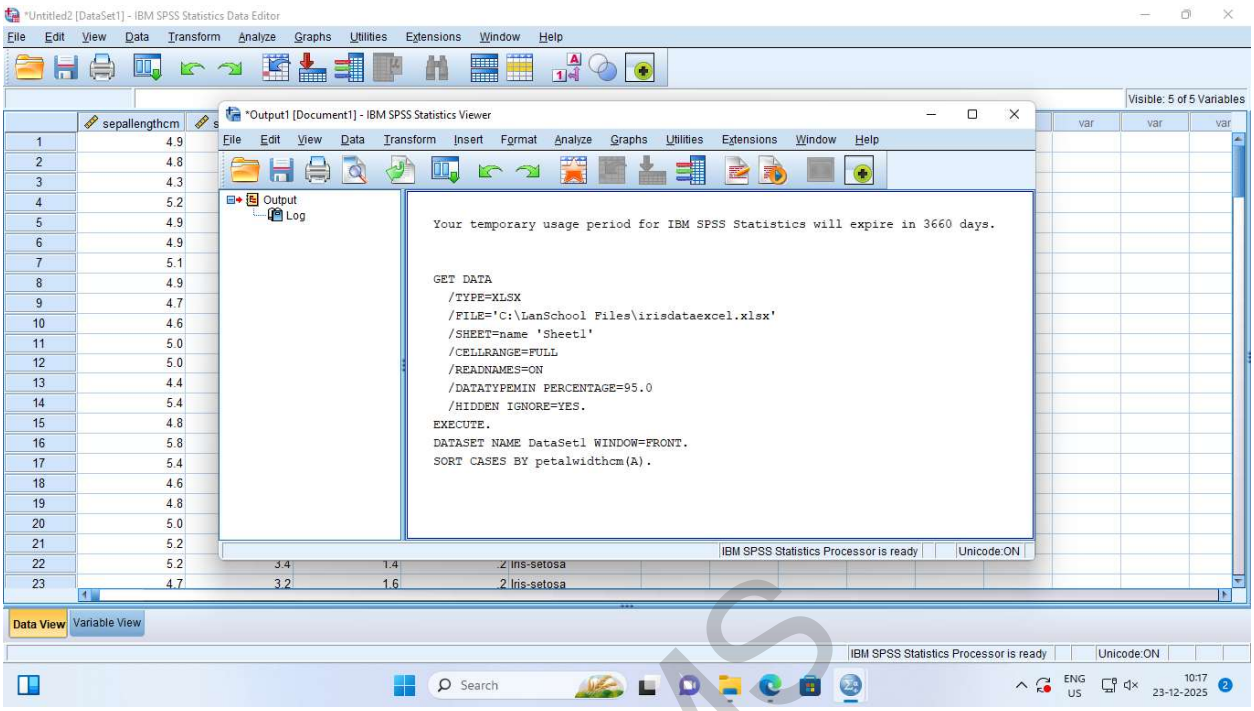

Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 0 of 0 Variables
Open Data
Look in: LanSchool Fil...
Hair Salon Client Data for Filtering and Sorting - answer.xlsx
IF EXERCISES.xlsx
irisdataexcel.xlsx
NEW EXERISE.xlsx
payroll macro exercise.xlsx
File name: irisdataexcel.xlsx
Files of type: Excel (*.xls, *.xlsx, *.xlsm)
Encoding:
Open
Paste
Cancel
Help
Retrieve File From Repository...
Data View
Variable View
Open data document
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:14
23-12-2025


Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 0 of 0 Variables
Read Excel File
C:\LanSchool Files\irisdataexcel.xlsx
Worksheet: Sheet1 [A1:E151]
Range:
Read variable names from first row of data
Percentage of values that determine data type: 95
Ignore hidden rows and columns
Remove leading spaces from string values
Remove trailing spaces from string values
Preview
sepalle... sepalw... petalle... petalwi...
1 5.1 3.5 1.4 0.2
2 4.9 3.0 1.4 0.2
3 4.7 3.2 1.3 0.2
4 4.6 3.1 1.5 0.2
5 5.0 3.6 1.4 0.2
6 5.4 3.9 1.7 0.4
7 4.6 3.4 1.4 0.3
Final data type is based on all data and can be different from the preview, which is based on the first 200 data rows. The preview displays only the first 500 columns.
OK Paste Reset Cancel Help
Data View Variable View
Open data document
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:14
23-12-2025


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 5 of 5 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target
1 5.1 3.5 1.4 .2 Iris-setosa
2 4.9 3.0 1.4 .2 Iris-setosa
3 4.7 3.2 1.3 .2 Iris-setosa
4 4.6 3.1 1.5 .2 Iris-setosa
5 5.0 3.6 1.4 .2 Iris-setosa
6 5.4 3.9 1.7 .4 Iris-setosa
7 4.6 3.4 1.4 .3 Iris-setosa
8 5.0 3.4 1.5 .2 Iris-setosa
9 4.4 2.9 1.4 .2 Iris-setosa
10 4.9 3.1 1.5 .1 Iris-setosa
11 5.4 3.7 1.5 .2 Iris-setosa
12 4.8 3.4 1.6 .2 Iris-setosa
13 4.8 3.0 1.4 .1 Iris-setosa
14 4.3 3.0 1.1 .1 Iris-setosa
15 5.8 4.0 1.2 .2 Iris-setosa
16 5.7 4.4 1.5 .4 Iris-setosa
17 5.4 3.9 1.3 .4 Iris-setosa
18 5.1 3.5 1.4 .3 Iris-setosa
19 5.7 3.8 1.7 .3 Iris-setosa
20 5.1 3.8 1.5 .3 Iris-setosa
21 5.4 3.4 1.7 .2 Iris-setosa
22 5.1 3.7 1.5 .4 Iris-setosa
23 4.6 3.6 1.0 .2 Iris-setosa
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:15
23-12-2025


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define date and time...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Adjust String Widths Across Files
Merge Files
Restructure...
Rake Weights...
Propensity Score Matching...
Case Control Matching...
Aggregate...
Orthogonal Design
Split into Files
Copy Dataset
Split File...
Select Cases...
Weight Cases...
10 : petalwidthcm
Visible: 5 of 5 Variables
petalwidthcm
target
Data View
Variable View
Sort Cases...
IBM SPSS Statistics Processor is ready
Unicode:ON


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
10 : petalwidthcm
Visible: 5 of 5 Variables
sepallengthcm
sepalwidthcm
petallengthcm
petalwidthcm
target
Sort Cases
Sort by:
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
target
petal width (cm) [pet...
Sort Order
Ascending
Descending
Save Sorted Data
Save file with sorted data
File...
Create an index
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

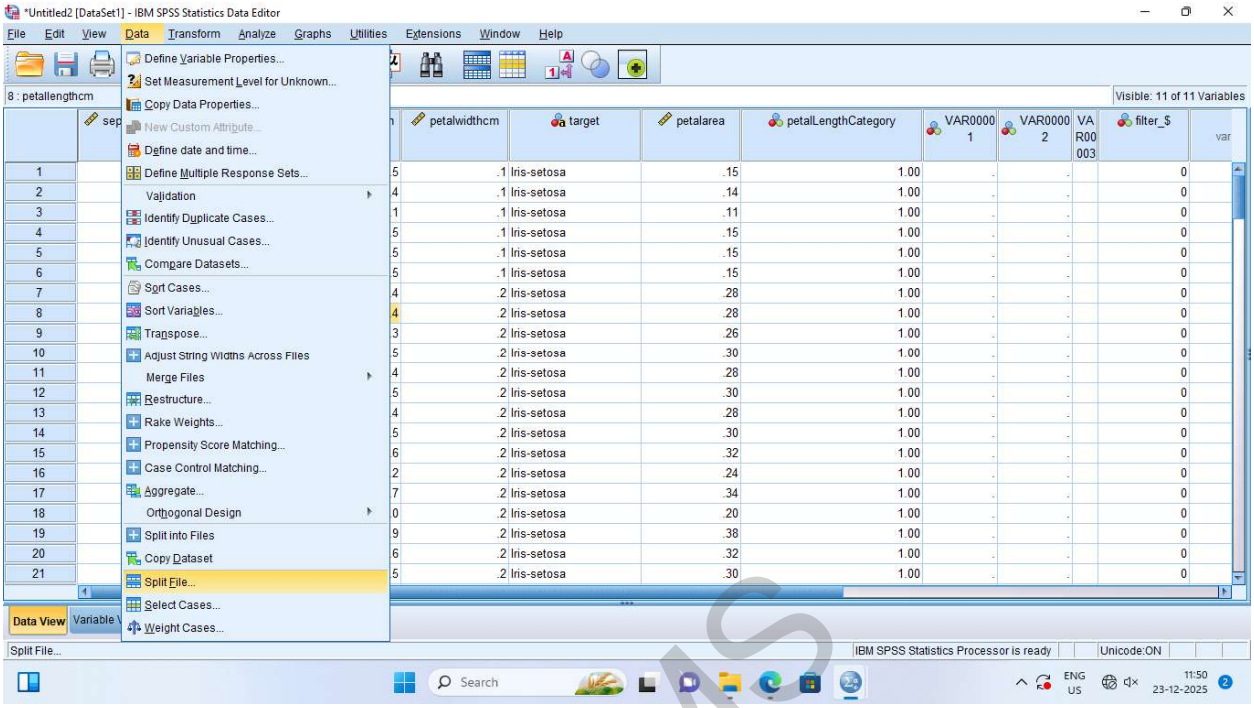

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 5 of 5 Variables
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
Your temporary usage period for IBM SPSS Statistics will expire in 3660 days.
GET DATA
/TYPE=XLSX
/FILE='C:\LanSchool Files\irisdataexcel.xlsx'
/SHEET=name 'Sheet1'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
SORT CASES BY petalwidthcm(A).
IBM SPSS Statistics Processor is ready
Unicode:ON
sepallengthcm
Data View
Variable View
Search
ENG
US
10:17
23-12-2025

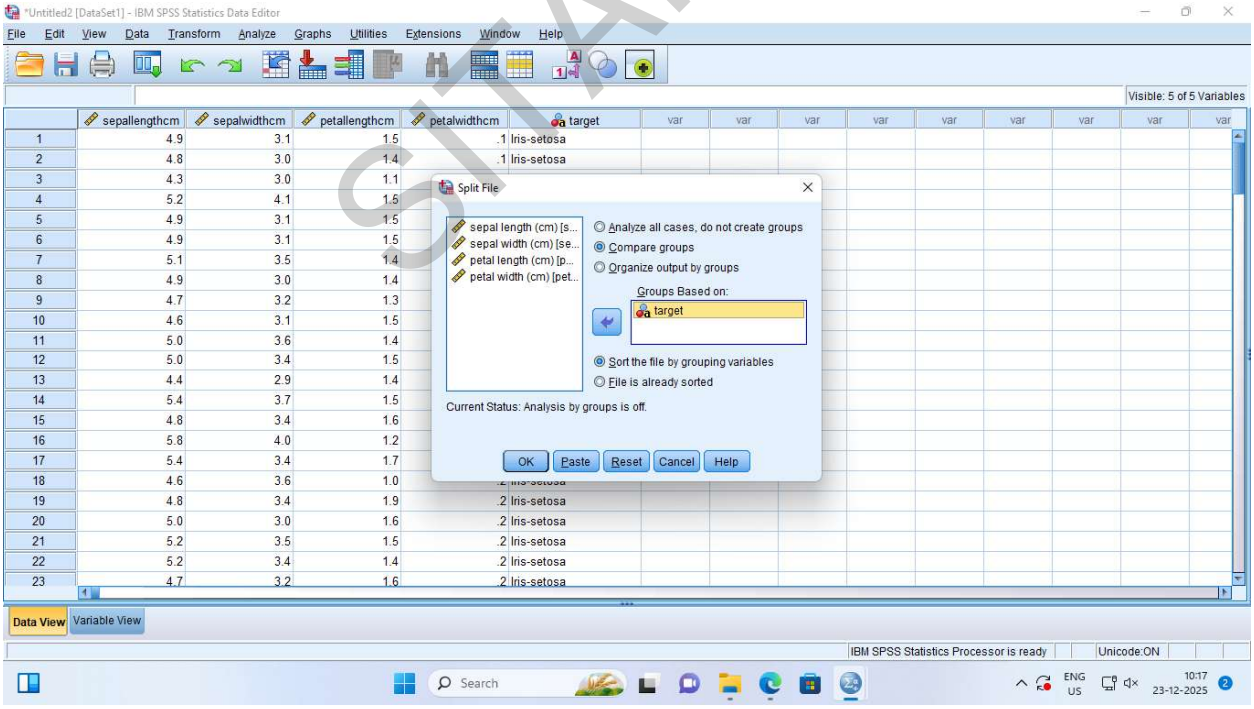
1.B)Split File

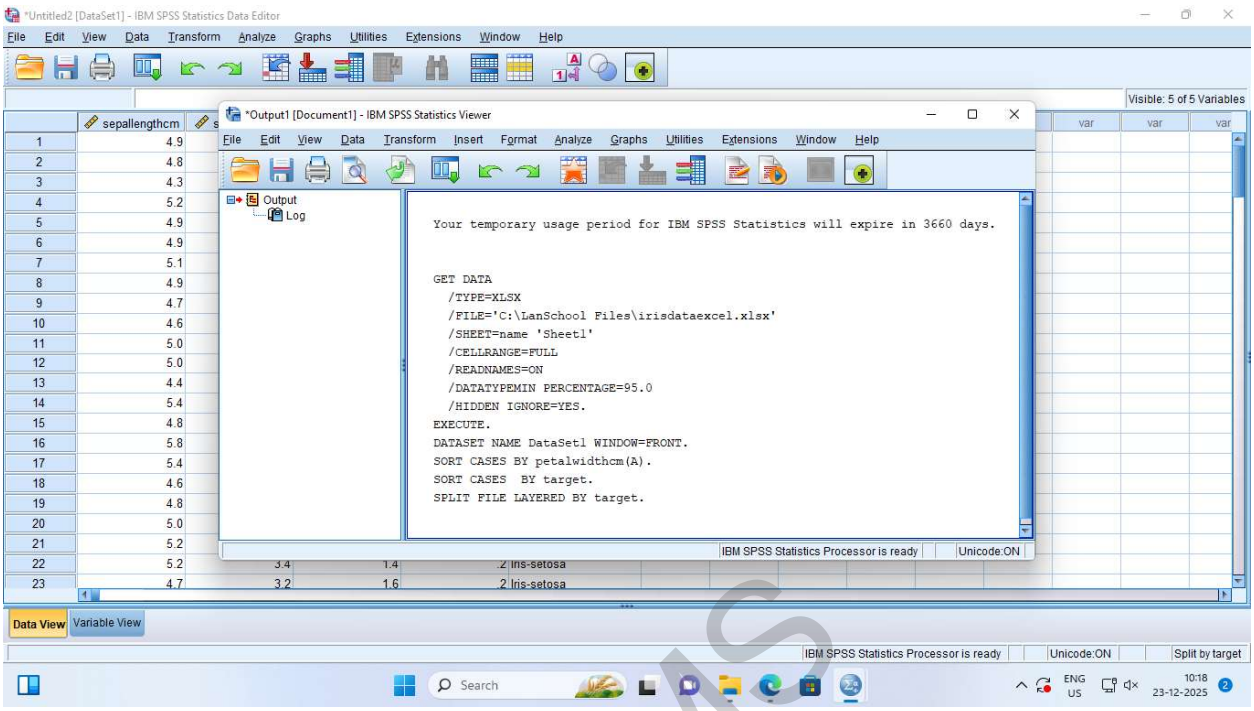
Aim: To perform analyses separately for subgroups defined by one or more categorical variables without physically splitting the data file.

Procedure Step by Step:

1. Open your dataset.
2. First, sort the data by the grouping variable: **Data > Sort Cases**, select the grouping variable, and click OK. (Recommended for organized output.)
3. Go to **Data > Split File**.
4. In the Split File dialog box, select **Analyze all cases, do not create groups** if currently split (to turn off), or choose:
    - **Compare groups** (output in single table comparing groups) or
    - **Organize output by groups** (separate tables for each group).
5. Move the grouping variable(s) to the "Groups Based on" box.
6. Click **OK**.
7. Subsequent analyses (e.g., frequencies, descriptives) will be performed separately for each group.
8. To turn off Split File: Repeat steps 3-4 and select **Analyze all cases, do not create groups**, then OK.


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define date and time...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Adjust String Widths Across Files
Merge Files
Restructure...
Rake Weights...
Propensity Score Matching...
Case Control Matching...
Aggregate...
Orthogonal Design
Split into Files
Copy Dataset
Split File...
Select Cases...
Weight Cases...
8 : petallengthcm
Visible: 11 of 11 Variables
petalwidthcm
target
petalarea
petalLengthCategory
VAR00001
VAR00002
VAR00003
filter_$
Iris-setosa
Data View
Variable View
Split File...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
11:50
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 5 of 5 Variables
sepallengthcm
sepalwidthcm
petallengthcm
petalwidthcm
target
Iris-setosa
Split File
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
petal width (cm) [pet...
Analyze all cases, do not create groups
Compare groups
Organize output by groups
Groups Based on:
target
Sort the file by grouping variables
File is already sorted
Current Status: Analysis by groups is off.
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:17
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 5 of 5 Variables
sepallengthcm
var
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
Your temporary usage period for IBM SPSS Statistics will expire in 3660 days.
GET DATA
/TYPE=XLSX
/FILE='C:\LanSchool Files\irisdataexcel.xlsx'
/SHEET=name 'Sheet1'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
SORT CASES BY petalwidthcm(A).
SORT CASES BY target.
SPLIT FILE LAYERED BY target.
IBM SPSS Statistics Processor is ready
Unicode:ON
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG
US
10:18
23-12-2025

1.C) Compute Variable

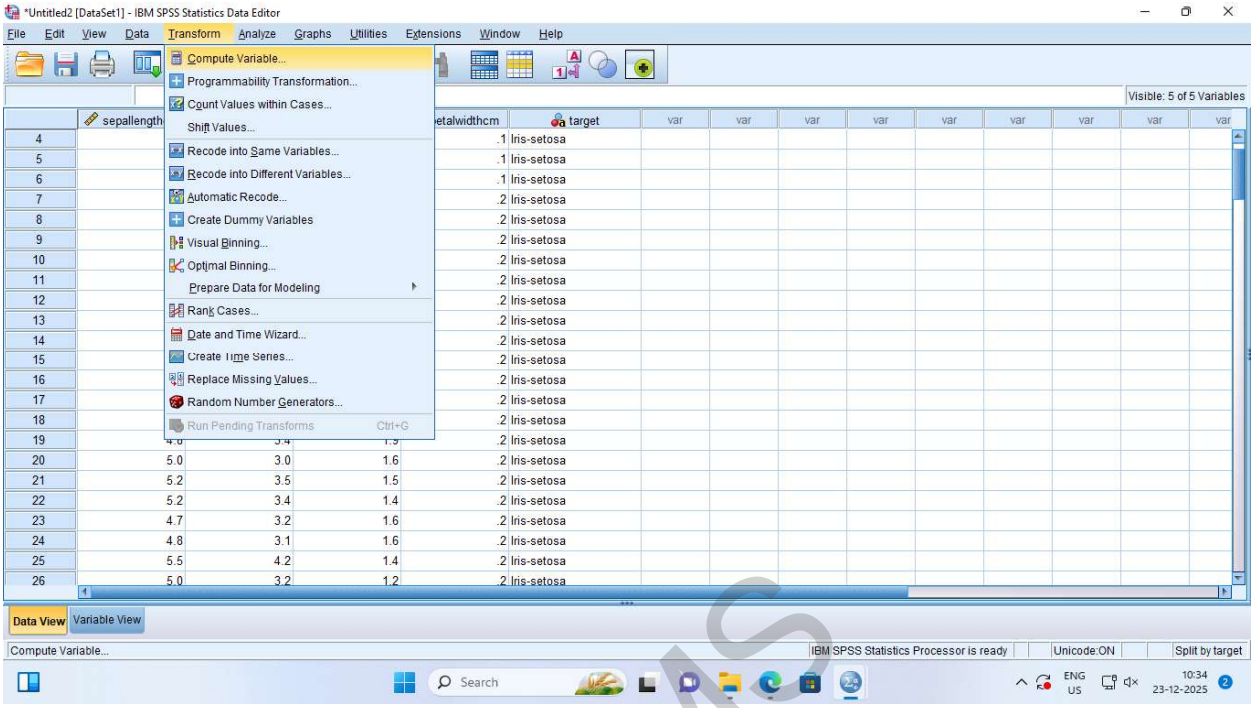
Aim: To create a new variable based on computations (e.g., arithmetic, functions) from existing variables.

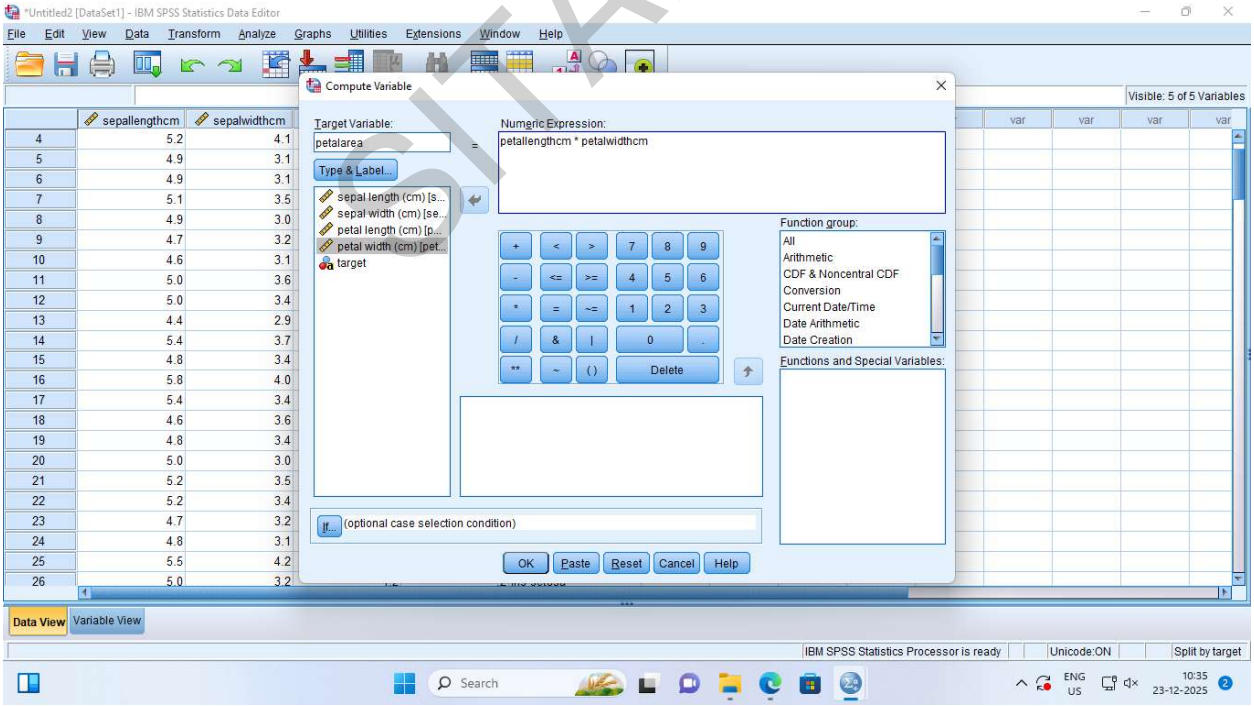
Procedure Step by Step:

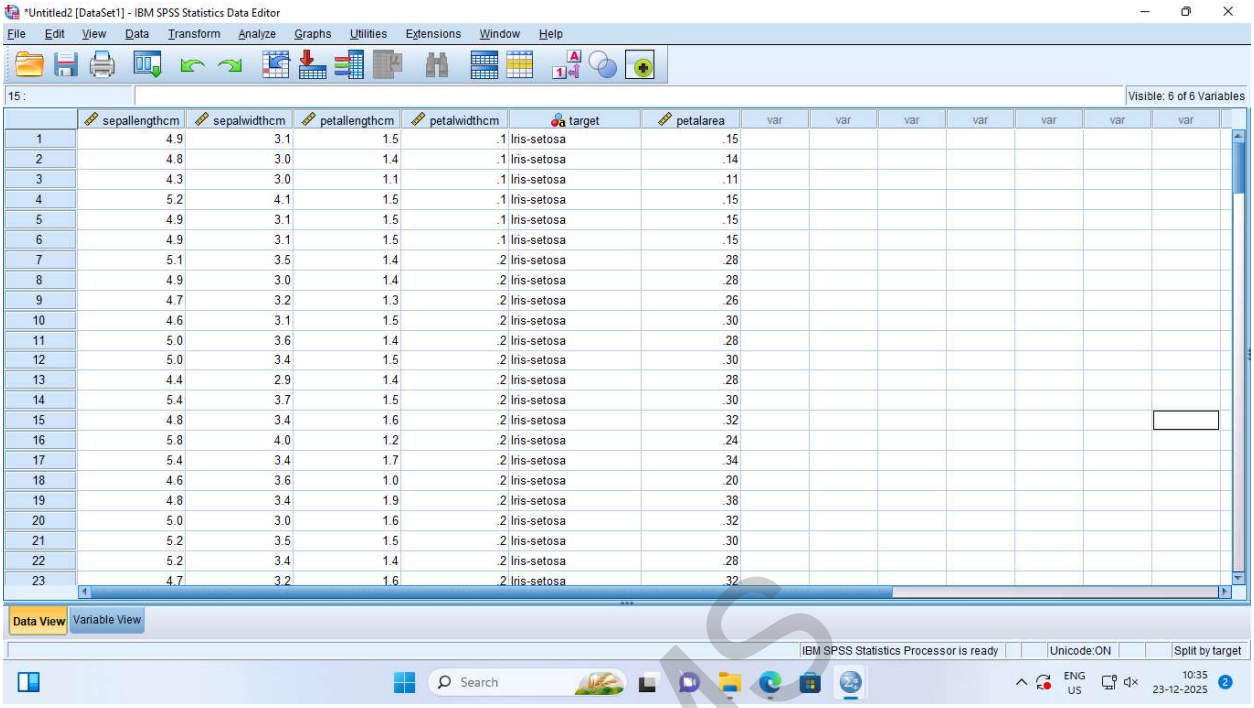
1. Open your dataset.
2. Go to **Transform > Compute Variable**.
3. In the Compute Variable dialog box, enter the name of the new **Target Variable** (must be a new name).
4. In the **Numeric Expression** box, build the formula:
    - Use the keypad or type directly.
    - Include existing variables (move from left list), operators (+, -, *, /, ** for power), and functions (e.g., MEAN, SUM, from the Functions list).

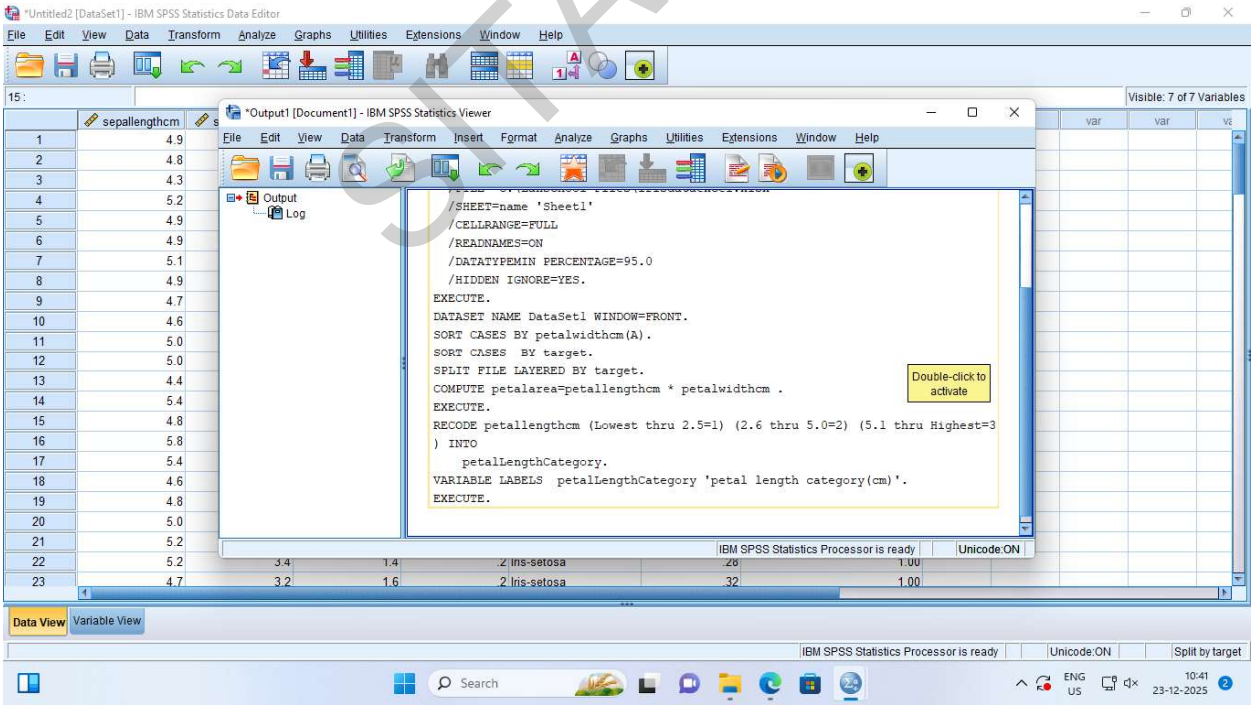
    Example: For average of three scores: MEAN(var1, var2, var3)

5. Optionally, click **Type & Label** to set label or type for the new variable.
6. Click **OK**.
7. The new variable appears as a new column in Data View (at the end).
8. Check results in Data View or run descriptives.


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Compute Variable...
Programmability Transformation...
Count Values within Cases...
Shift Values...
Recode into Same Variables...
Recode into Different Variables...
Automatic Recode...
Create Dummy Variables
Visual Binning...
Optimal Binning...
Prepare Data for Modeling
Rank Cases...
Date and Time Wizard...
Create Time Series...
Replace Missing Values...
Random Number Generators...
Run Pending Transforms Ctrl+G
Visible: 5 of 5 Variables
target
Iris-setosa
Data View
Variable View
Compute Variable...
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:34
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Compute Variable
Target Variable:
petalarea
Type & Label...
Numeric Expression:
petallengthcm * petalwidthcm
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
petal width (cm) [pet...
target
Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation
Functions and Special Variables:
Delete
If... (optional case selection condition)
OK Paste Reset Cancel Help
Visible: 5 of 5 Variables
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:35
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
15 :
Visible: 6 of 6 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea var var var var var var var
1 4.9 3.1 1.5 .1 Iris-setosa .15
2 4.8 3.0 1.4 .1 Iris-setosa .14
3 4.3 3.0 1.1 .1 Iris-setosa .11
4 5.2 4.1 1.5 .1 Iris-setosa .15
5 4.9 3.1 1.5 .1 Iris-setosa .15
6 4.9 3.1 1.5 .1 Iris-setosa .15
7 5.1 3.5 1.4 .2 Iris-setosa .28
8 4.9 3.0 1.4 .2 Iris-setosa .28
9 4.7 3.2 1.3 .2 Iris-setosa .26
10 4.6 3.1 1.5 .2 Iris-setosa .30
11 5.0 3.6 1.4 .2 Iris-setosa .28
12 5.0 3.4 1.5 .2 Iris-setosa .30
13 4.4 2.9 1.4 .2 Iris-setosa .28
14 5.4 3.7 1.5 .2 Iris-setosa .30
15 4.8 3.4 1.6 .2 Iris-setosa .32
16 5.8 4.0 1.2 .2 Iris-setosa .24
17 5.4 3.4 1.7 .2 Iris-setosa .34
18 4.6 3.6 1.0 .2 Iris-setosa .20
19 4.8 3.4 1.9 .2 Iris-setosa .38
20 5.0 3.0 1.6 .2 Iris-setosa .32
21 5.2 3.5 1.5 .2 Iris-setosa .30
22 5.2 3.4 1.4 .2 Iris-setosa .28
23 4.7 3.2 1.6 .2 Iris-setosa .32
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:35
23-12-2025

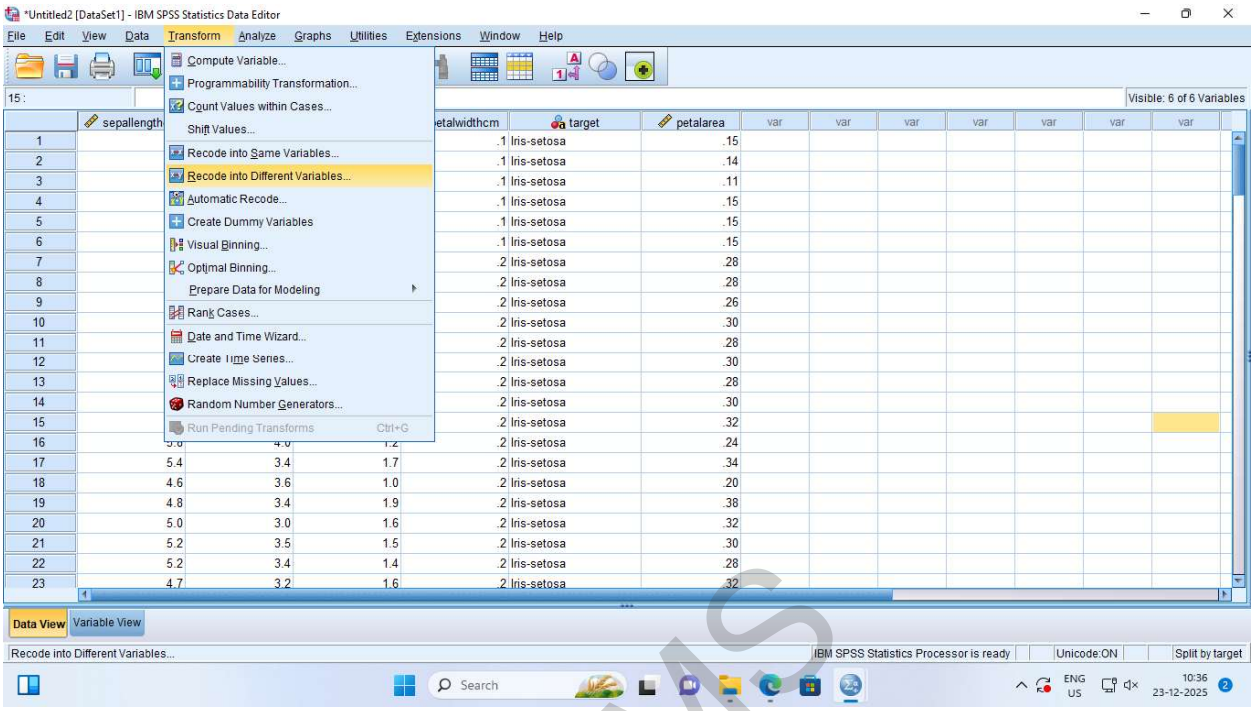

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
15 :
Visible: 7 of 7 Variables
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
/SHEET=name 'Sheet1'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
SORT CASES BY petalwidthcm(A).
SORT CASES BY target.
SPLIT FILE LAYERED BY target.
COMPUTE petalarea=petallengthcm * petalwidthcm .
EXECUTE.
RECODE petallengthcm (Lowest thru 2.5=1) (2.6 thru 5.0=2) (5.1 thru Highest=3
) INTO
petalLengthCategory.
VARIABLE LABELS petalLengthCategory 'petal length category(cm)'.
EXECUTE.
Double-click to activate
IBM SPSS Statistics Processor is ready
Unicode:ON
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:41
23-12-2025

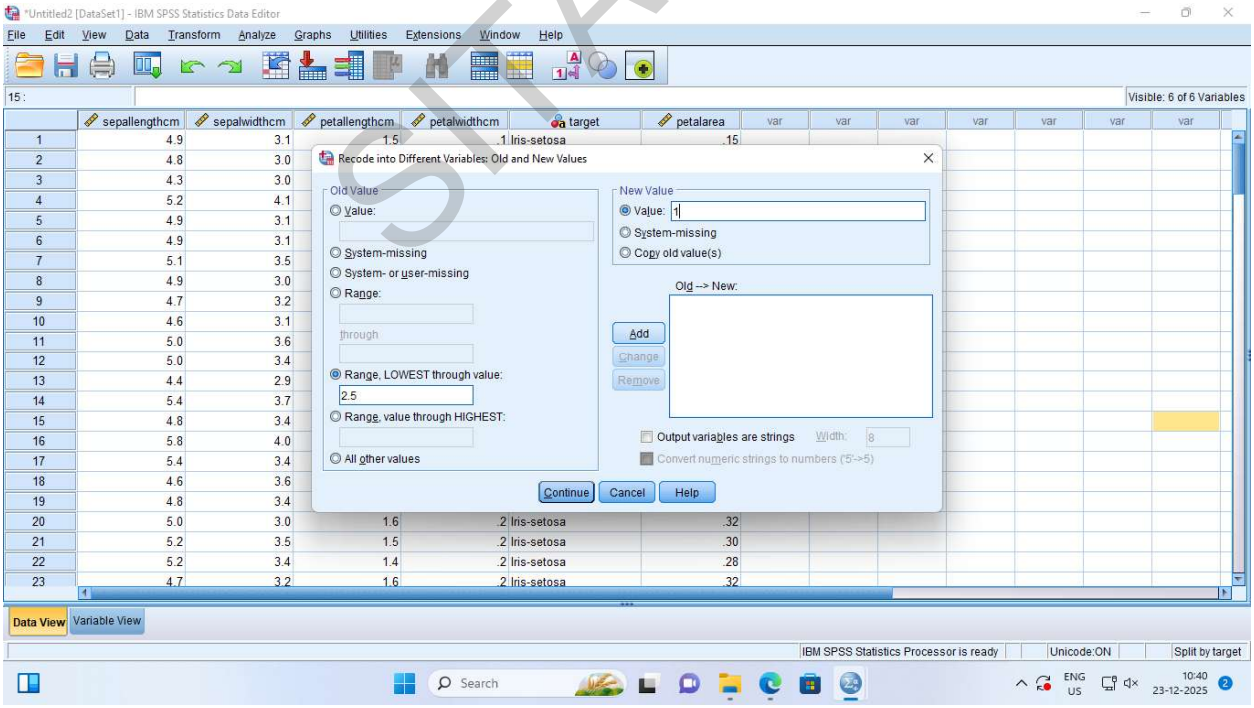
1.D) Recode Variable

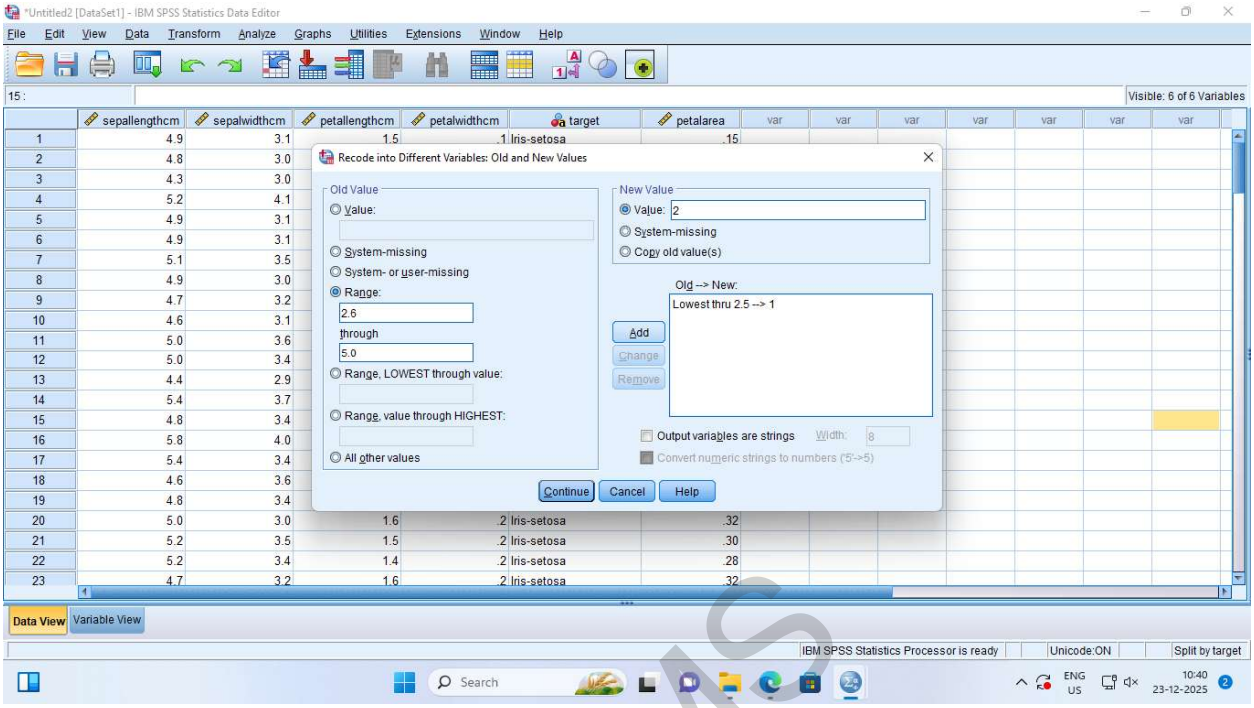
Aim: To recode values of an existing variable into new values, creating a new variable (recommended to preserve original data).

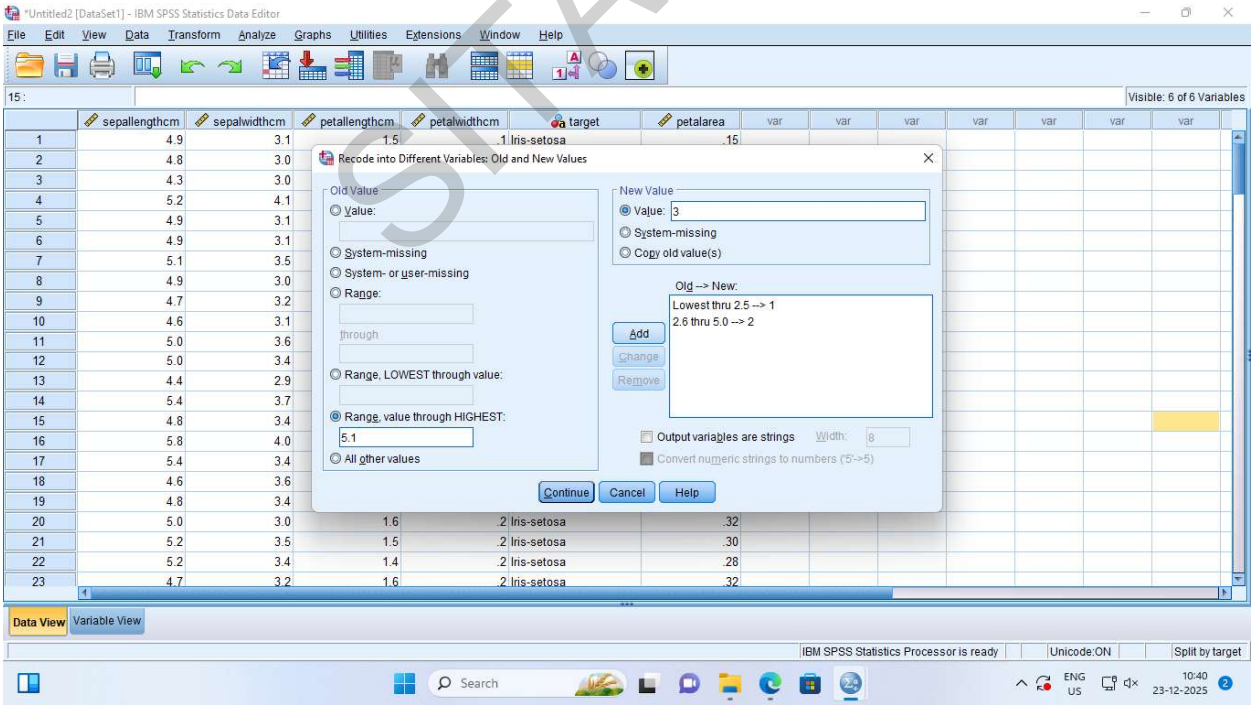
Procedure Step by Step (Recode into Different Variables):

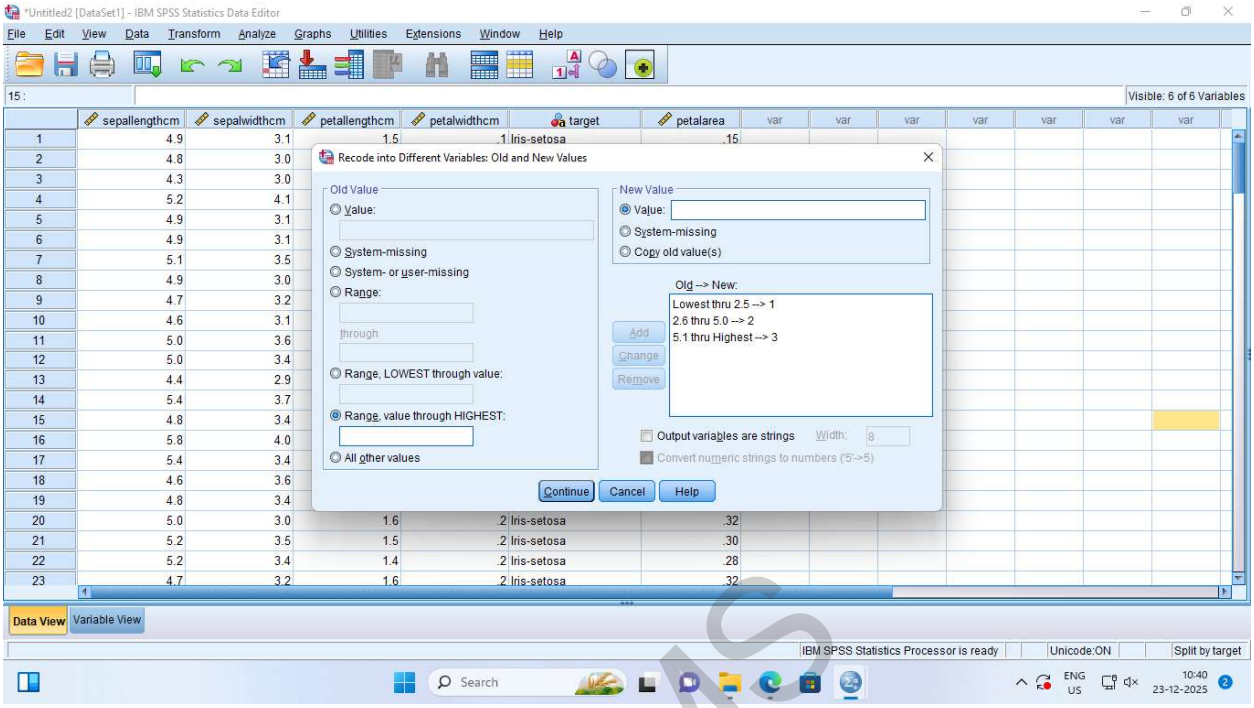
1. Open your dataset.
2. Go to **Transform > Recode into Different Variables**.
3. Move the input variable(s) to the "Input Variable -> Output Variable" box.
4. For each, enter the **Output Variable** name and optionally label, then click **Change**.
5. Click **Old and New Values**.
6. In the Old and New Values dialog:
    - Specify **Old Value** (e.g., single value, Range, All other values) and **New Value**.
    - Click **Add** for each mapping.
    - Example: Old 1 -> New 0; Old 2 thru 5 -> New 1.
7. Optionally, handle System-missing or copy old values.
8. Click **Continue**, then **OK**.
9. The new recoded variable appears in Data View.
10. Add value labels in Variable View if needed.

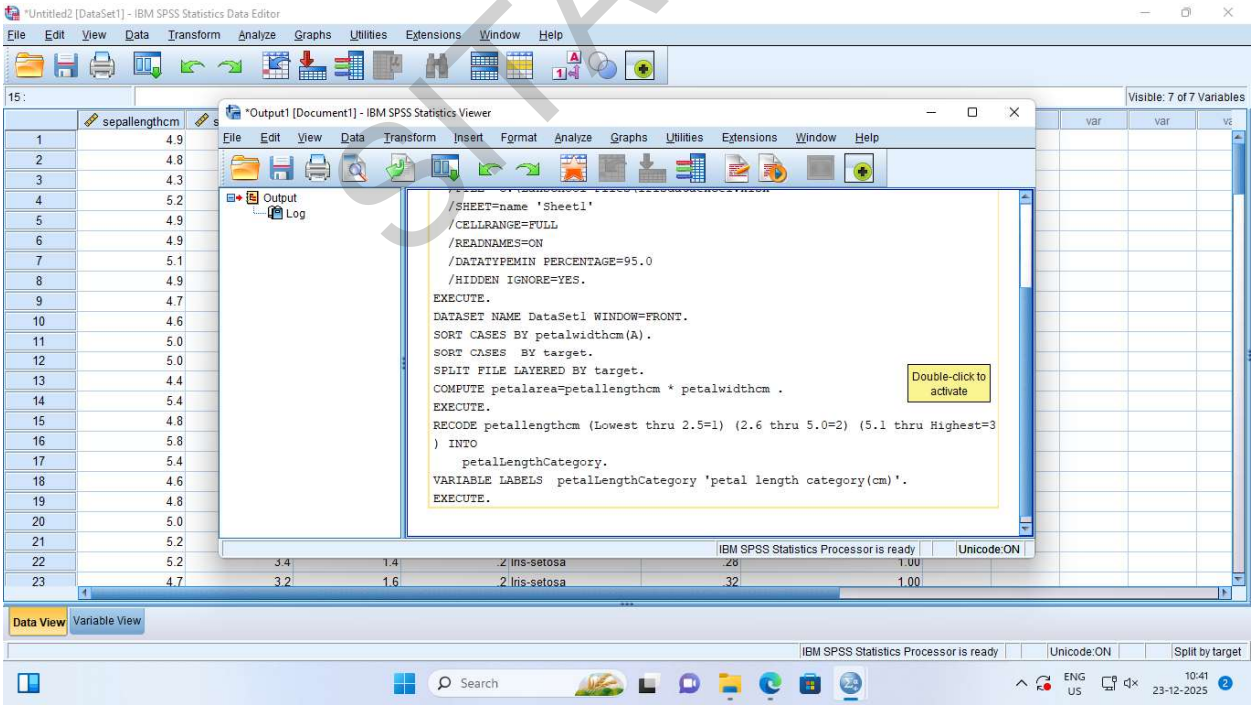

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Compute Variable...
Programmability Transformation...
Count Values within Cases...
Shift Values...
Recode into Same Variables...
Recode into Different Variables...
Automatic Recode...
Create Dummy Variables
Visual Binning...
Optimal Binning...
Prepare Data for Modeling
Rank Cases...
Date and Time Wizard...
Create Time Series...
Replace Missing Values...
Random Number Generators...
Run Pending Transforms Ctrl+G
Visible: 6 of 6 Variables
Data View
Variable View
Recode into Different Variables...
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

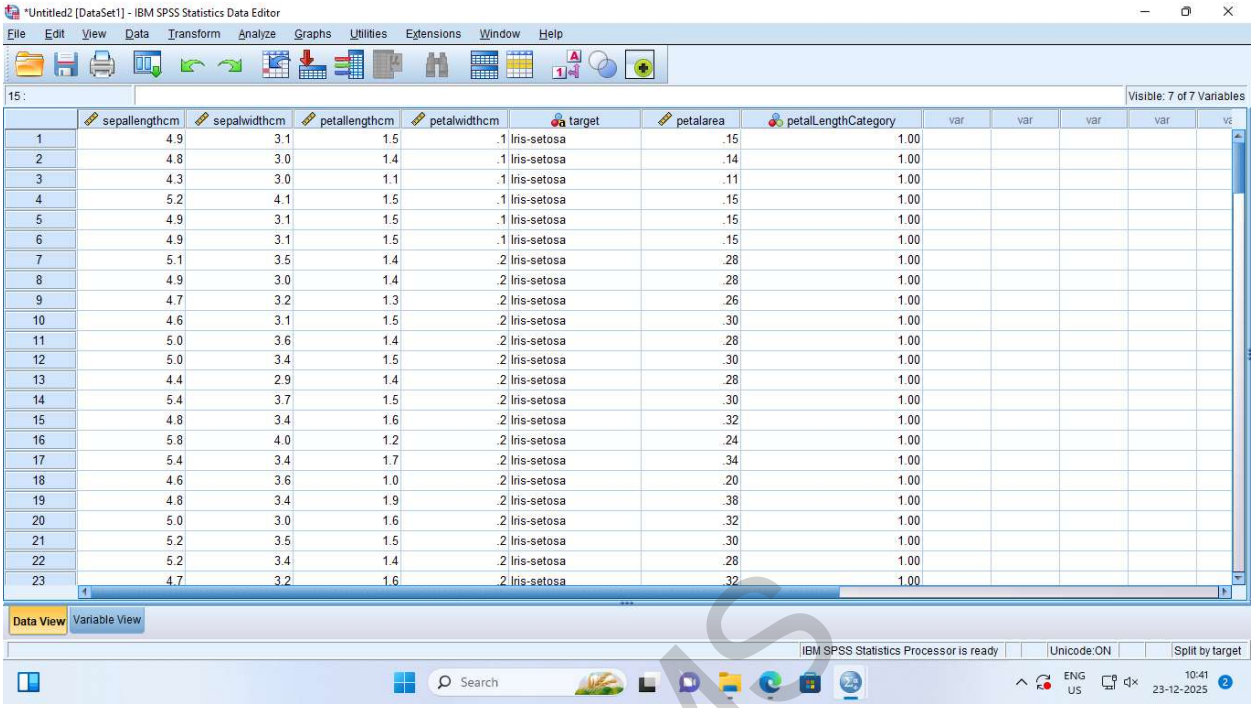

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range, LOWEST through value:
2.5
Range, value through HIGHEST:
All other values
New Value
Value: 1
System-missing
Copy old value(s)
Old --> New:
Add
Change
Remove
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

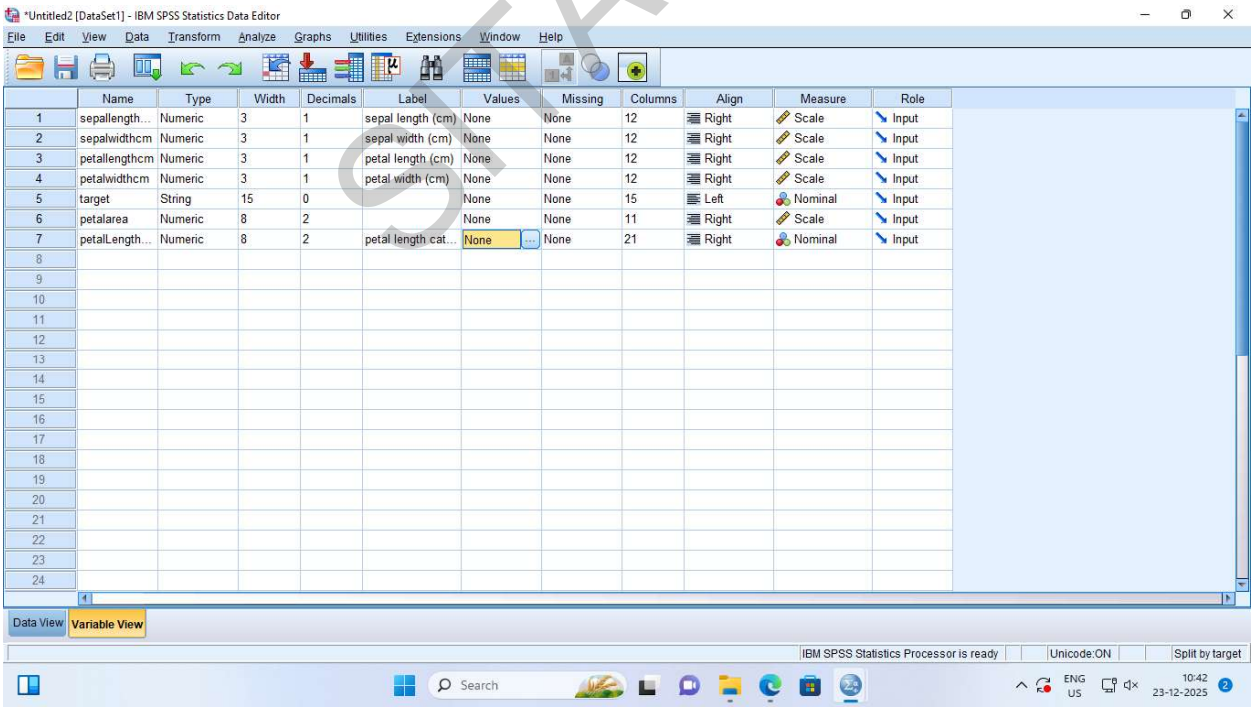

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
2.6
through
5.0
Range, LOWEST through value:
Range, value through HIGHEST:
All other values
New Value
Value: 2
System-missing
Copy old value(s)
Old --> New:
Lowest thru 2.5 --> 1
Add
Change
Remove
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

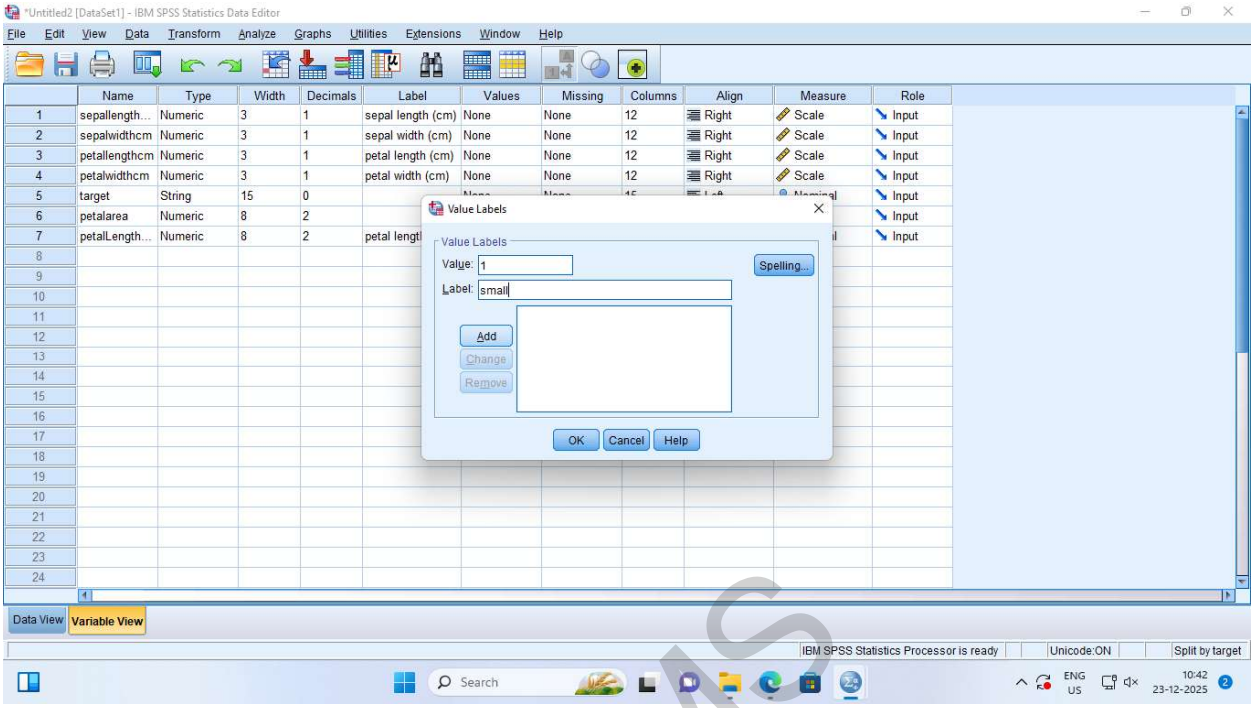

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range, LOWEST through value:
Range, value through HIGHEST:
5.1
All other values
New Value
Value: 3
System-missing
Copy old value(s)
Old --> New:
Lowest thru 2.5 --> 1
2.6 thru 5.0 --> 2
Add
Change
Remove
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

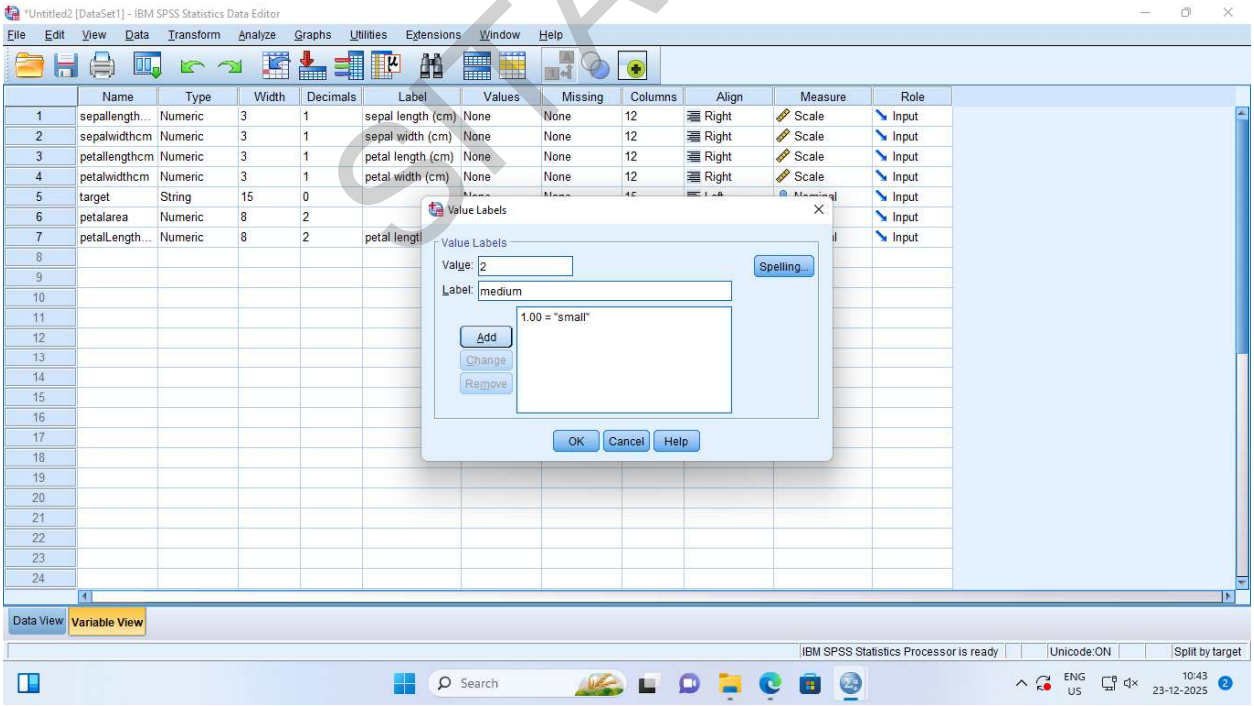

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 6 of 6 Variables
Recode into Different Variables: Old and New Values
Old Value
Value:
System-missing
System- or user-missing
Range:
through
Range, LOWEST through value:
Range, value through HIGHEST:
All other values
New Value
Value:
System-missing
Copy old value(s)
Old --> New:
Lowest thru 2.5 --> 1
2.6 thru 5.0 --> 2
5.1 thru Highest --> 3
Add
Change
Remove
Output variables are strings
Width:
Convert numeric strings to numbers ('5'->5)
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

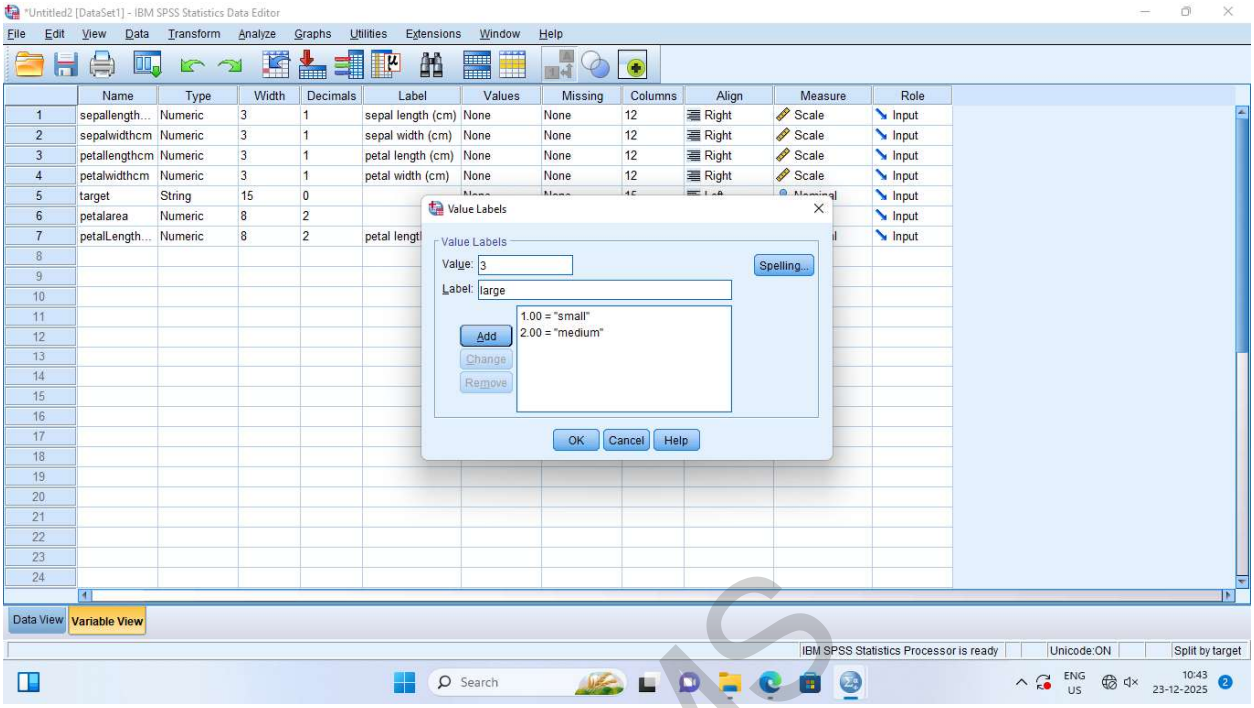

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 7 of 7 Variables
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
/SHEET=name 'Sheet1'
/CELLRANGE=FULL
/READNAMES=ON
/DATATYPEMIN PERCENTAGE=95.0
/HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
SORT CASES BY petalwidthcm(A).
SORT CASES BY target.
SPLIT FILE LAYERED BY target.
COMPUTE petalarea=petallengthcm * petalwidthcm .
EXECUTE.
RECODE petallengthcm (Lowest thru 2.5=1) (2.6 thru 5.0=2) (5.1 thru Highest=3
) INTO
petalLengthCategory.
VARIABLE LABELS petalLengthCategory 'petal length category(cm)'.
EXECUTE.
Double-click to activate
IBM SPSS Statistics Processor is ready
Unicode:ON
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target

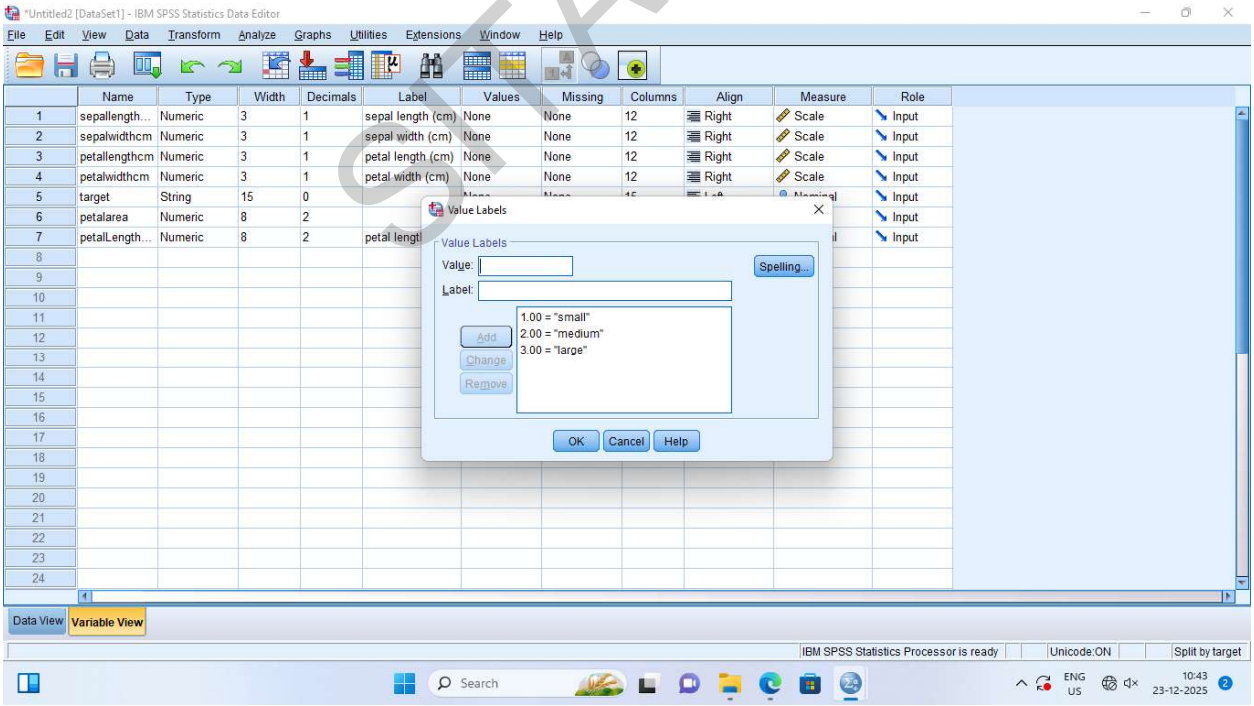

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
15 :
Visible: 7 of 7 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory var var var var var
1 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
2 4.8 3.0 1.4 .1 Iris-setosa .14 1.00
3 4.3 3.0 1.1 .1 Iris-setosa .11 1.00
4 5.2 4.1 1.5 .1 Iris-setosa .15 1.00
5 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
6 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
7 5.1 3.5 1.4 .2 Iris-setosa .28 1.00
8 4.9 3.0 1.4 .2 Iris-setosa .28 1.00
9 4.7 3.2 1.3 .2 Iris-setosa .26 1.00
10 4.6 3.1 1.5 .2 Iris-setosa .30 1.00
11 5.0 3.6 1.4 .2 Iris-setosa .28 1.00
12 5.0 3.4 1.5 .2 Iris-setosa .30 1.00
13 4.4 2.9 1.4 .2 Iris-setosa .28 1.00
14 5.4 3.7 1.5 .2 Iris-setosa .30 1.00
15 4.8 3.4 1.6 .2 Iris-setosa .32 1.00
16 5.8 4.0 1.2 .2 Iris-setosa .24 1.00
17 5.4 3.4 1.7 .2 Iris-setosa .34 1.00
18 4.6 3.6 1.0 .2 Iris-setosa .20 1.00
19 4.8 3.4 1.9 .2 Iris-setosa .38 1.00
20 5.0 3.0 1.6 .2 Iris-setosa .32 1.00
21 5.2 3.5 1.5 .2 Iris-setosa .30 1.00
22 5.2 3.4 1.4 .2 Iris-setosa .28 1.00
23 4.7 3.2 1.6 .2 Iris-setosa .32 1.00
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:41
23-12-2025

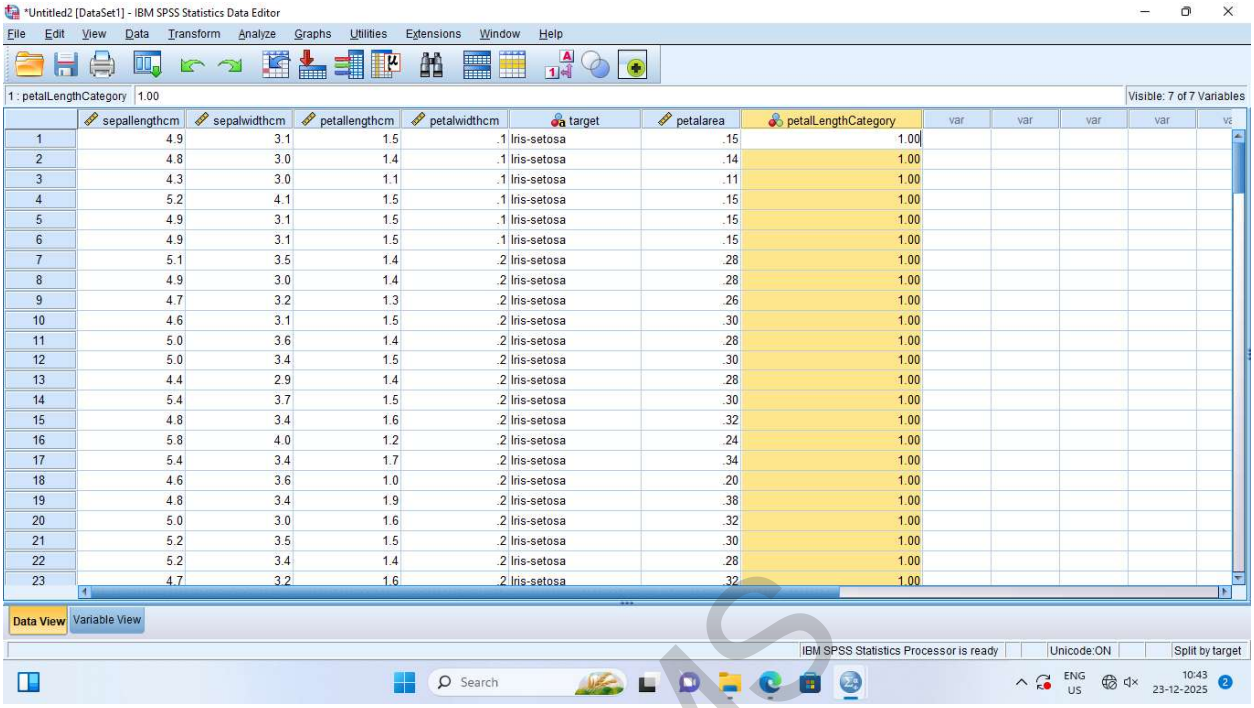

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sepallength... Numeric 3 1 sepal length (cm) None None 12 Right Scale Input
2 sepalwidthcm Numeric 3 1 sepal width (cm) None None 12 Right Scale Input
3 petallengthcm Numeric 3 1 petal length (cm) None None 12 Right Scale Input
4 petalwidthcm Numeric 3 1 petal width (cm) None None 12 Right Scale Input
5 target String 15 0 None None 15 Left Nominal Input
6 petalarea Numeric 8 2 None None 11 Right Scale Input
7 petalLength... Numeric 8 2 petal length cat... None None 21 Right Nominal Input
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:42
23-12-2025

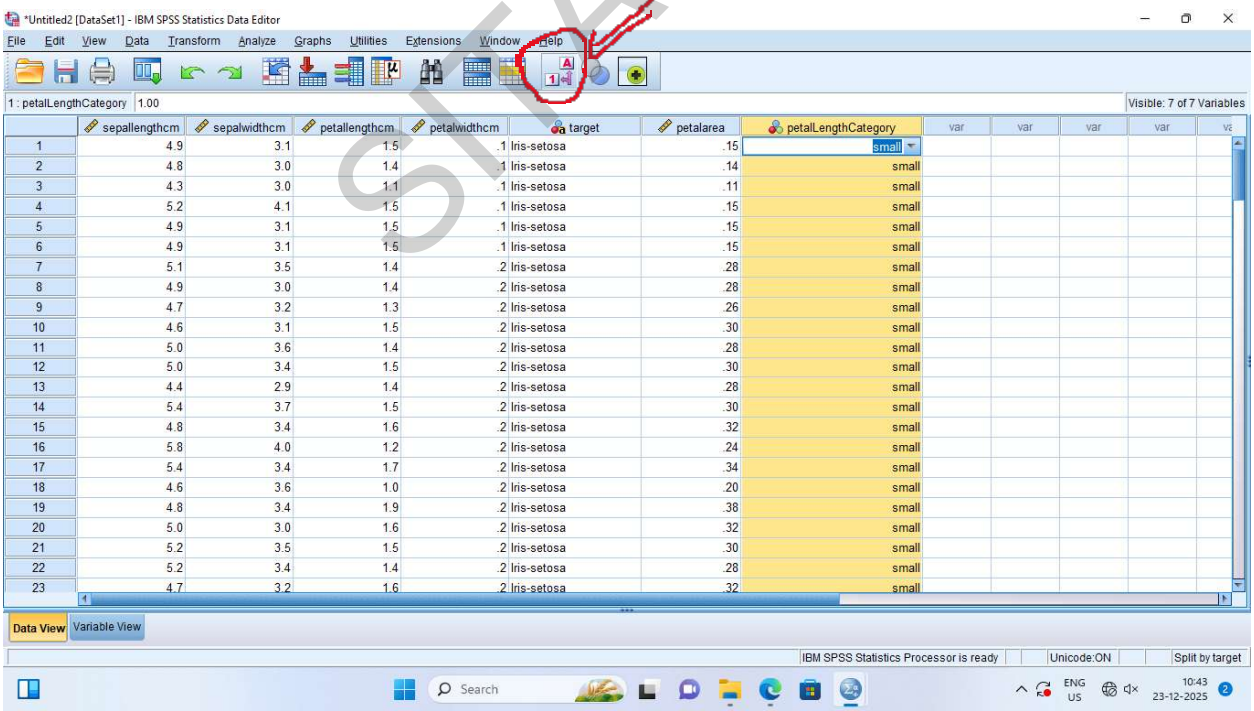

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sepallength... Numeric 3 1 sepal length (cm) None None 12 Right Scale Input
2 sepalwidthcm Numeric 3 1 sepal width (cm) None None 12 Right Scale Input
3 petallengthcm Numeric 3 1 petal length (cm) None None 12 Right Scale Input
4 petalwidthcm Numeric 3 1 petal width (cm) None None 12 Right Scale Input
5 target String 15 0 Input
6 petalarea Numeric 8 2 Input
7 petalLength... Numeric 8 2 petal lengtl Input
Value Labels
Value Labels
Value: 1
Label: small
Spelling...
Add
Change
Remove
OK Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:42
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sepallength... Numeric 3 1 sepal length (cm) None None 12 Right Scale Input
2 sepalwidthcm Numeric 3 1 sepal width (cm) None None 12 Right Scale Input
3 petallengthcm Numeric 3 1 petal length (cm) None None 12 Right Scale Input
4 petalwidthcm Numeric 3 1 petal width (cm) None None 12 Right Scale Input
5 target String 15 0 Input
6 petalarea Numeric 8 2 Input
7 petalLength... Numeric 8 2 petal lengtl Input
Value Labels
Value Labels
Value: 2
Label: medium
1.00 = "small"
Spelling...
Add
Change
Remove
OK Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:43
23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sepallength... Numeric 3 1 sepal length (cm) None None 12 Right Scale Input
2 sepalwidthcm Numeric 3 1 sepal width (cm) None None 12 Right Scale Input
3 petallengthcm Numeric 3 1 petal length (cm) None None 12 Right Scale Input
4 petalwidthcm Numeric 3 1 petal width (cm) None None 12 Right Scale Input
5 target String 15 0 Input
6 petalarea Numeric 8 2 Input
7 petalLength... Numeric 8 2 petal lengtl Input
Value Labels
Value Labels
Value: 3
Label: large
Spelling...
Add
Change
Remove
1.00 = "small"
2.00 = "medium"
OK Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:43 23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 sepallength... Numeric 3 1 sepal length (cm) None None 12 Right Scale Input
2 sepalwidthcm Numeric 3 1 sepal width (cm) None None 12 Right Scale Input
3 petallengthcm Numeric 3 1 petal length (cm) None None 12 Right Scale Input
4 petalwidthcm Numeric 3 1 petal width (cm) None None 12 Right Scale Input
5 target String 15 0 Input
6 petalarea Numeric 8 2 Input
7 petalLength... Numeric 8 2 petal lengtl Input
Value Labels
Value Labels
Value:
Label:
Spelling...
Add
Change
Remove
1.00 = "small"
2.00 = "medium"
3.00 = "large"
OK Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:43 23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
1 : petalLengthCategory 1.00
Visible: 7 of 7 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory var var var var va
1 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
2 4.8 3.0 1.4 .1 Iris-setosa .14 1.00
3 4.3 3.0 1.1 .1 Iris-setosa .11 1.00
4 5.2 4.1 1.5 .1 Iris-setosa .15 1.00
5 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
6 4.9 3.1 1.5 .1 Iris-setosa .15 1.00
7 5.1 3.5 1.4 .2 Iris-setosa .28 1.00
8 4.9 3.0 1.4 .2 Iris-setosa .28 1.00
9 4.7 3.2 1.3 .2 Iris-setosa .26 1.00
10 4.6 3.1 1.5 .2 Iris-setosa .30 1.00
11 5.0 3.6 1.4 .2 Iris-setosa .28 1.00
12 5.0 3.4 1.5 .2 Iris-setosa .30 1.00
13 4.4 2.9 1.4 .2 Iris-setosa .28 1.00
14 5.4 3.7 1.5 .2 Iris-setosa .30 1.00
15 4.8 3.4 1.6 .2 Iris-setosa .32 1.00
16 5.8 4.0 1.2 .2 Iris-setosa .24 1.00
17 5.4 3.4 1.7 .2 Iris-setosa .34 1.00
18 4.6 3.6 1.0 .2 Iris-setosa .20 1.00
19 4.8 3.4 1.9 .2 Iris-setosa .38 1.00
20 5.0 3.0 1.6 .2 Iris-setosa .32 1.00
21 5.2 3.5 1.5 .2 Iris-setosa .30 1.00
22 5.2 3.4 1.4 .2 Iris-setosa .28 1.00
23 4.7 3.2 1.6 .2 Iris-setosa .32 1.00
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:43 23-12-2025

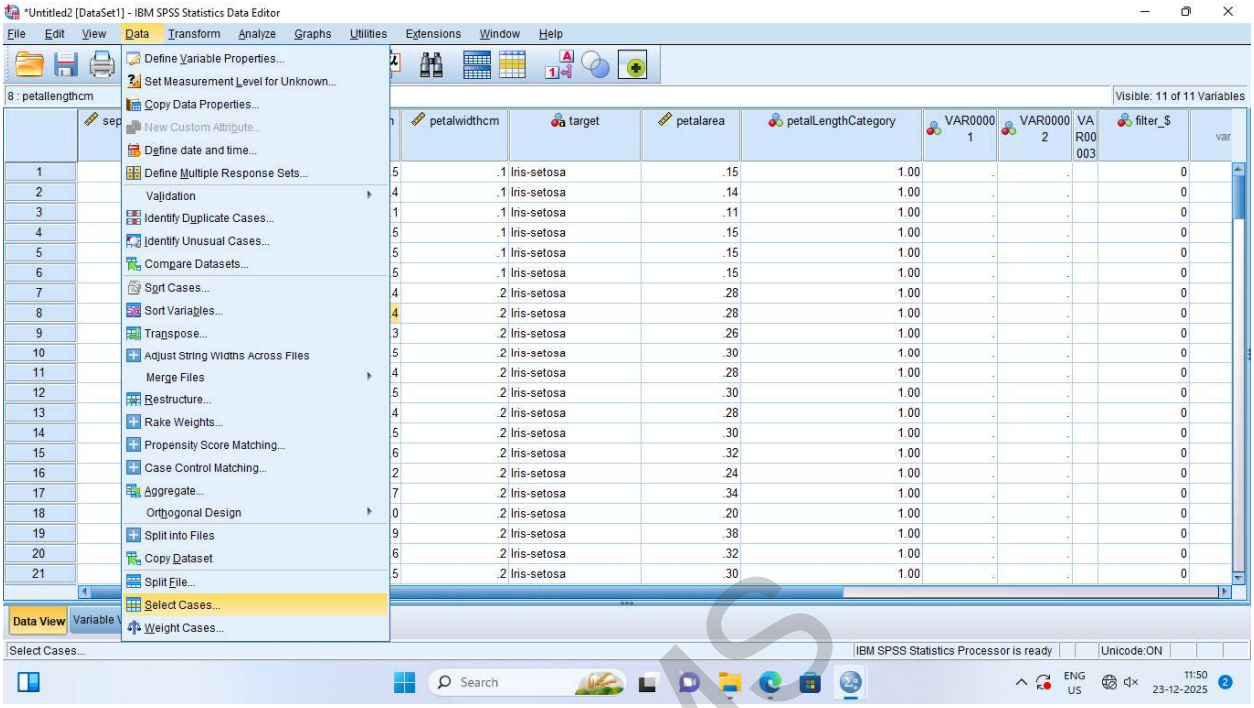

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
1 : petalLengthCategory 1.00
Visible: 7 of 7 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory var var var var va
1 4.9 3.1 1.5 .1 Iris-setosa .15 small
2 4.8 3.0 1.4 .1 Iris-setosa .14 small
3 4.3 3.0 1.1 .1 Iris-setosa .11 small
4 5.2 4.1 1.5 .1 Iris-setosa .15 small
5 4.9 3.1 1.5 .1 Iris-setosa .15 small
6 4.9 3.1 1.5 .1 Iris-setosa .15 small
7 5.1 3.5 1.4 .2 Iris-setosa .28 small
8 4.9 3.0 1.4 .2 Iris-setosa .28 small
9 4.7 3.2 1.3 .2 Iris-setosa .26 small
10 4.6 3.1 1.5 .2 Iris-setosa .30 small
11 5.0 3.6 1.4 .2 Iris-setosa .28 small
12 5.0 3.4 1.5 .2 Iris-setosa .30 small
13 4.4 2.9 1.4 .2 Iris-setosa .28 small
14 5.4 3.7 1.5 .2 Iris-setosa .30 small
15 4.8 3.4 1.6 .2 Iris-setosa .32 small
16 5.8 4.0 1.2 .2 Iris-setosa .24 small
17 5.4 3.4 1.7 .2 Iris-setosa .34 small
18 4.6 3.6 1.0 .2 Iris-setosa .20 small
19 4.8 3.4 1.9 .2 Iris-setosa .38 small
20 5.0 3.0 1.6 .2 Iris-setosa .32 small
21 5.2 3.5 1.5 .2 Iris-setosa .30 small
22 5.2 3.4 1.4 .2 Iris-setosa .28 small
23 4.7 3.2 1.6 .2 Iris-setosa .32 small
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Split by target
Search
ENG US
10:43 23-12-2025

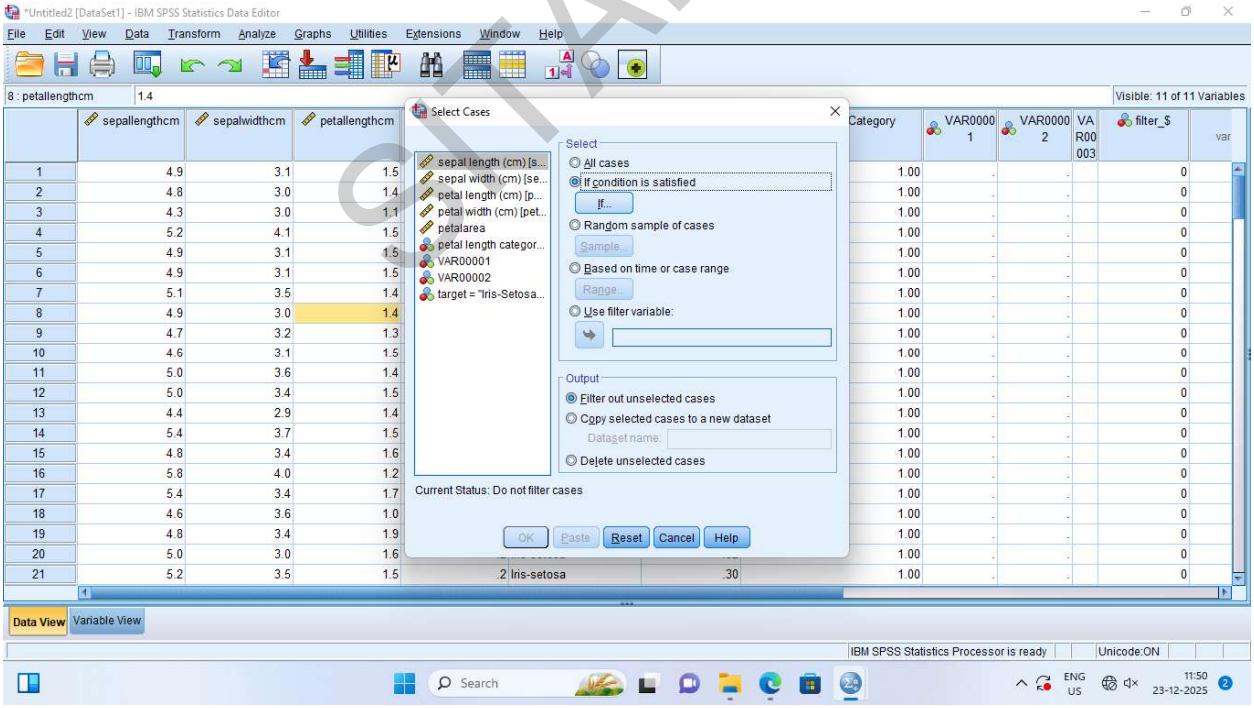
1.E) Select Cases

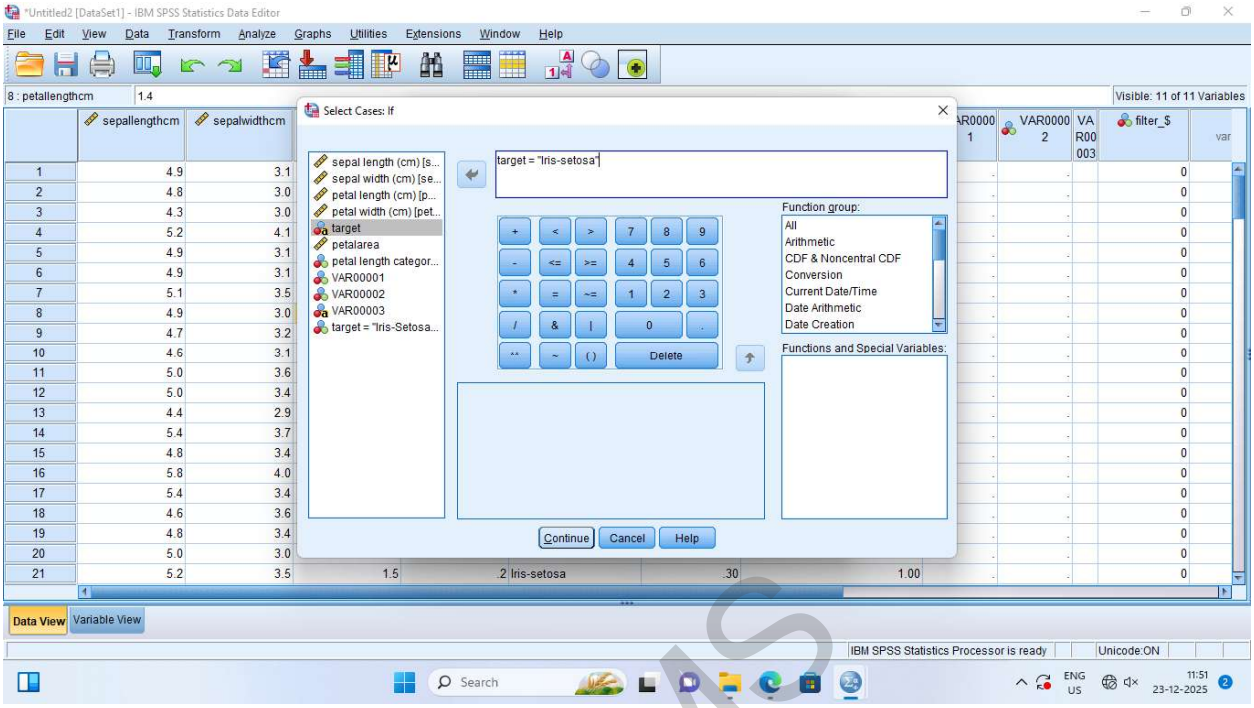
Aim: To temporarily select a subset of cases for analysis (e.g., based on conditions), filtering out others without deleting them.

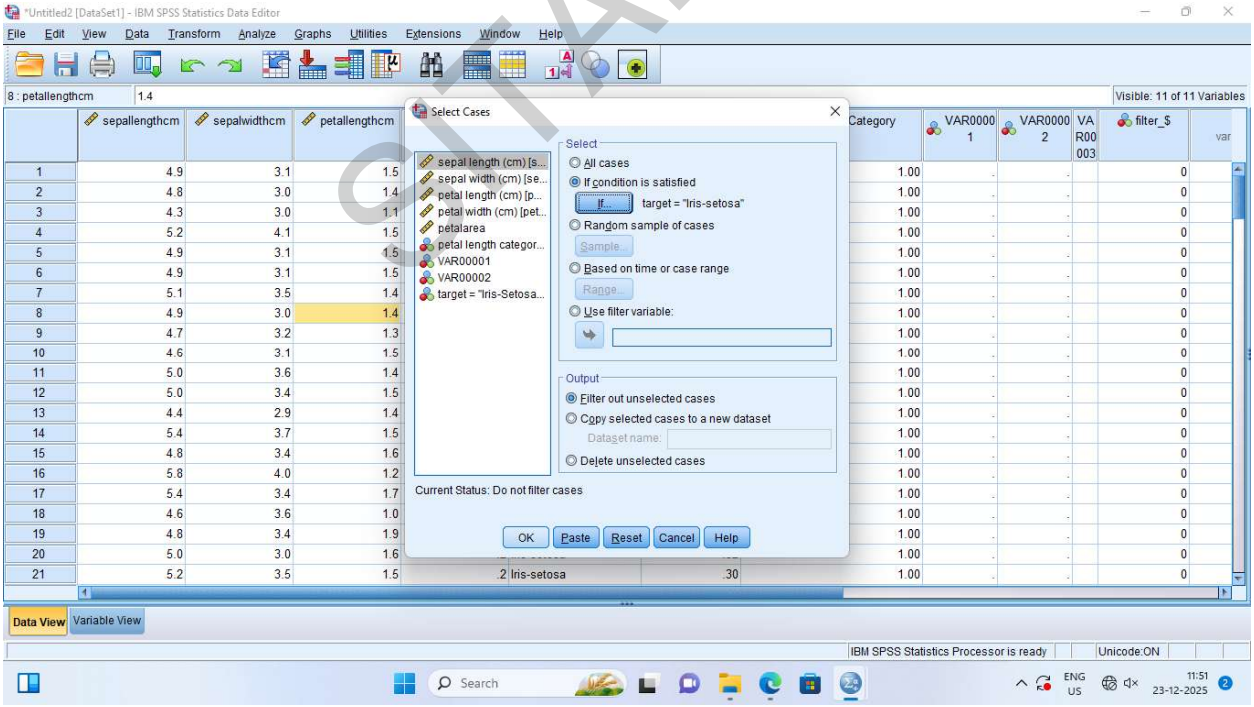
Procedure Step by Step:

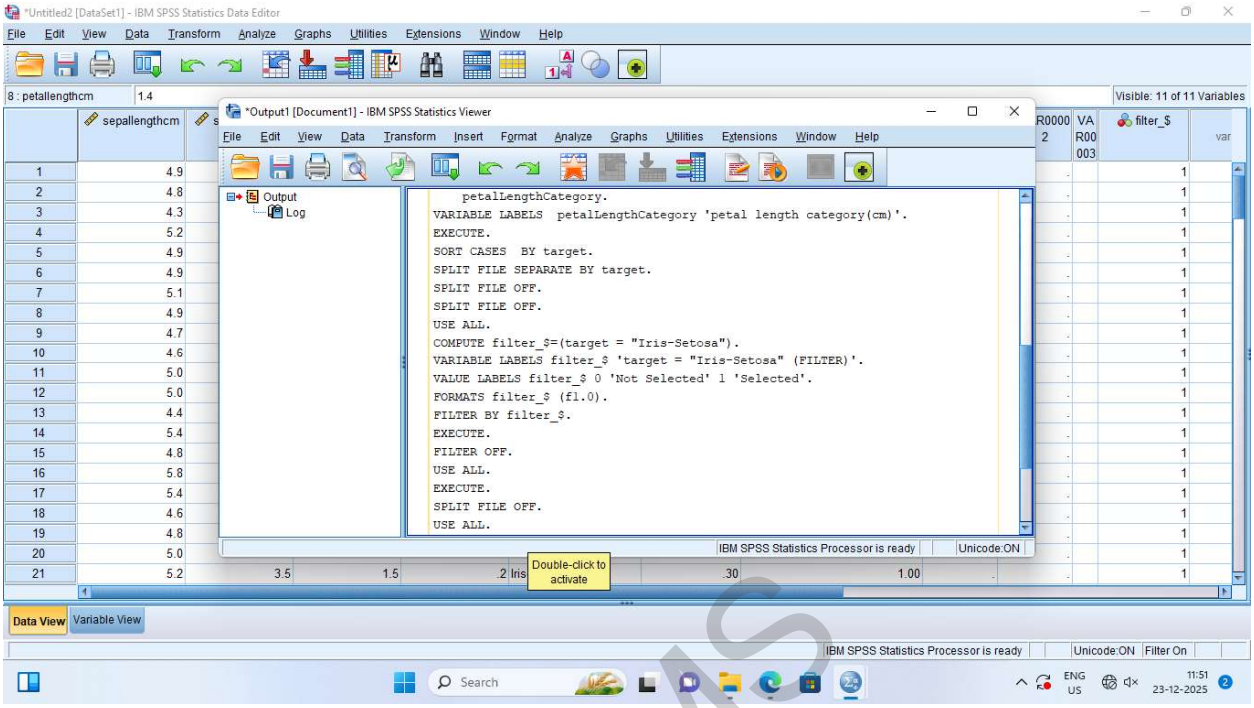
1. Open your dataset.
2. Go to **Data > Select Cases**.
3. Choose an option:
    - **All cases** (to reset).
    - **If condition is satisfied** (most common for conditional selection).
4. For conditional: Click **If**.
5. Build the expression (e.g., gender = 1 & age > 18). Use variables, operators (=, >, <, &, | for AND/OR).
6. Click **Continue**.
7. Choose output: **Filter** (creates filter variable, unselected cases hidden but kept) or **Delete** (permanently removes unselected).
8. Click **OK**.
9. A filter variable appears; "Filter On" shows in status bar. Analyses use only selected cases.
10. To reset: Repeat and select **All cases**, then OK.

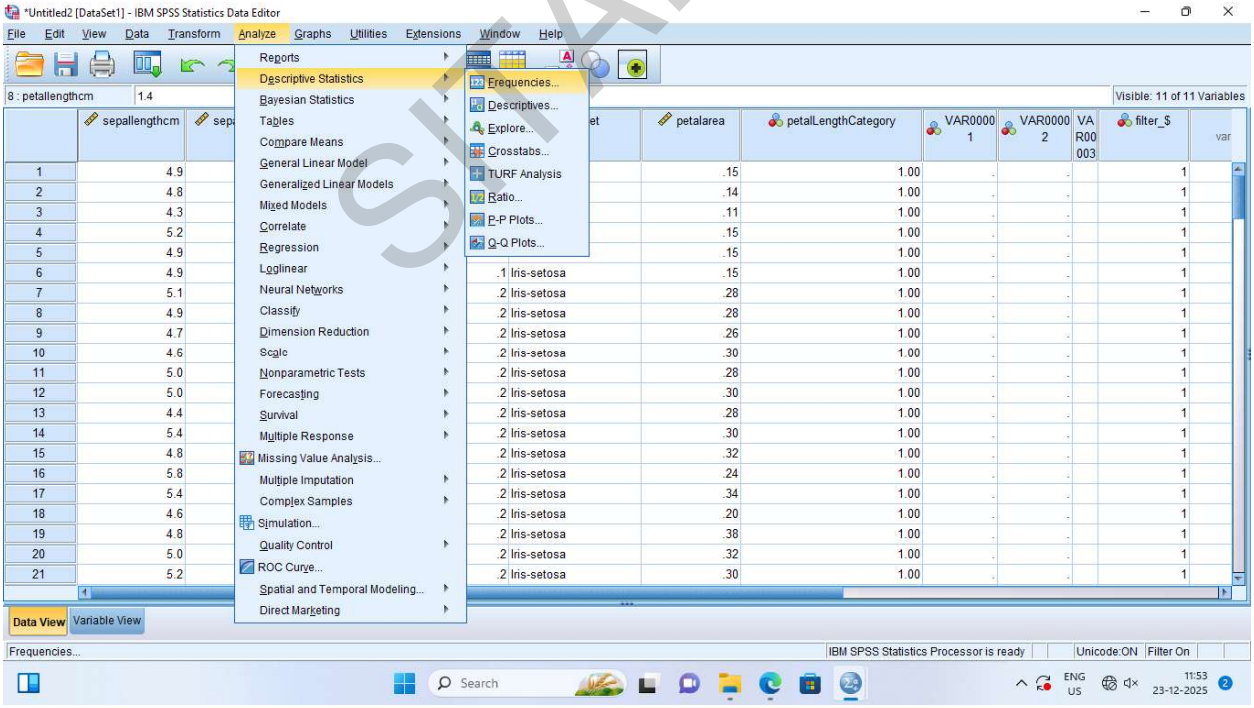

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define date and time...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Adjust String Widths Across Files
Merge Files
Restructure...
Rake Weights...
Propensity Score Matching...
Case Control Matching...
Aggregate...
Orthogonal Design
Split into Files
Copy Dataset
Split File...
Select Cases...
Weight Cases...
8 : petallengthcm
Visible: 11 of 11 Variables
petalwidthcm
target
petalarea
petalLengthCategory
VAR00001
VAR00002
VAR00003
filter_$
Data View
Variable View
Select Cases...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
11:50
23-12-2025

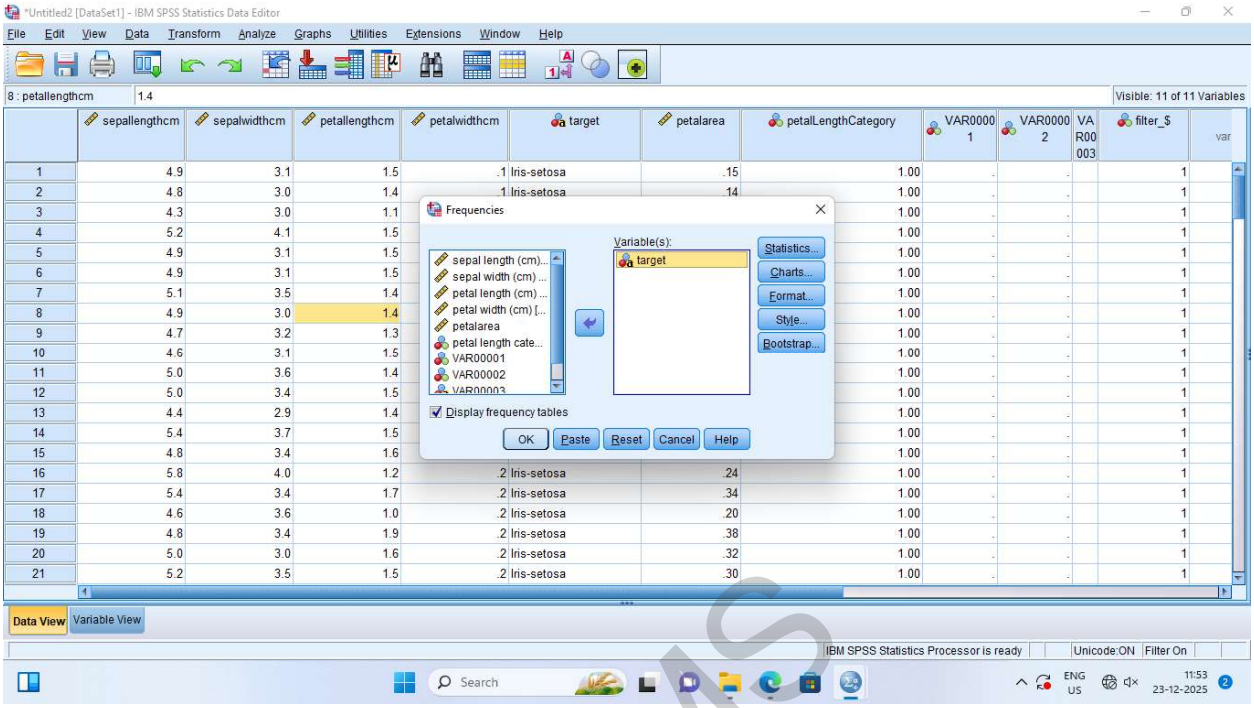

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm
1.4
Visible: 11 of 11 Variables
sepallengthcm
sepalwidthcm
petallengthcm
Select Cases
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
petal width (cm) [pet...
petalarea
petal length categor...
VAR00001
VAR00002
target = "Iris-Setosa...
Select
All cases
If condition is satisfied
If...
Random sample of cases
Sample...
Based on time or case range
Range...
Use filter variable:
Output
Filter out unselected cases
Copy selected cases to a new dataset
Dataset name:
Delete unselected cases
Current Status: Do not filter cases
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
11:50
23-12-2025

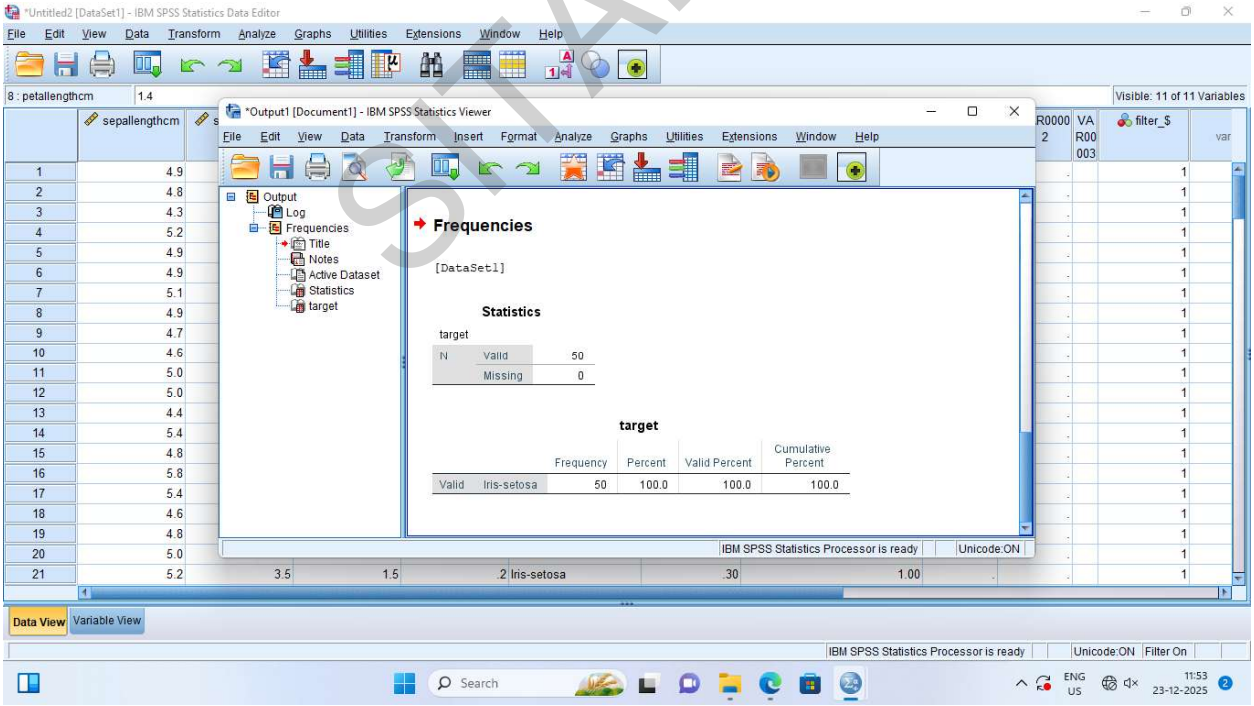

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
Select Cases: If
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
petal width (cm) [pet...
target
petalarea
petal length categor...
VAR00001
VAR00002
VAR00003
target = "Iris-Setosa...
target = "Iris-setosa"
Function group:
All
Arithmetic
CDF & Noncentral CDF
Conversion
Current Date/Time
Date Arithmetic
Date Creation
Functions and Special Variables:
Delete
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
11:51
23-12-2025

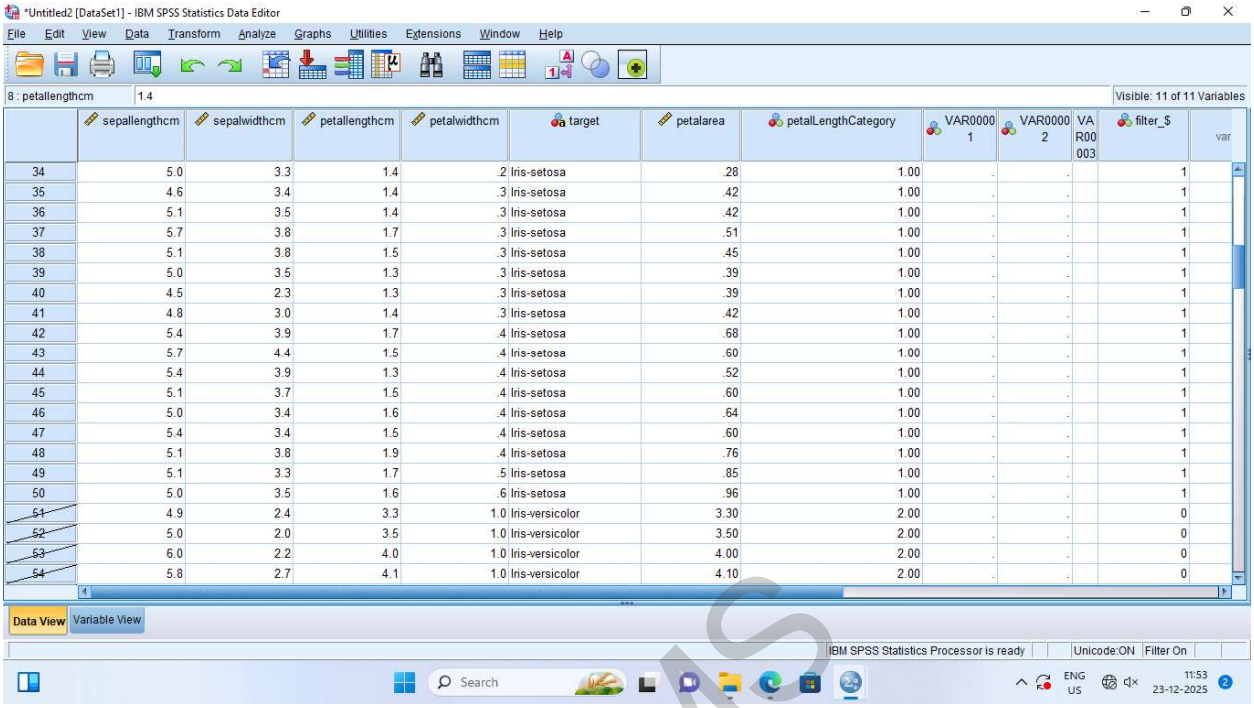

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
Select Cases
sepal length (cm) [s...
sepal width (cm) [se...
petal length (cm) [p...
petal width (cm) [pet...
petalarea
petal length categor...
VAR00001
VAR00002
target = "Iris-Setosa...
Select
All cases
If condition is satisfied
If...
target = "Iris-setosa"
Random sample of cases
Sample...
Based on time or case range
Range...
Use filter variable:
Output
Filter out unselected cases
Copy selected cases to a new dataset
Dataset name:
Delete unselected cases
Current Status: Do not filter cases
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
11:51
23-12-2025

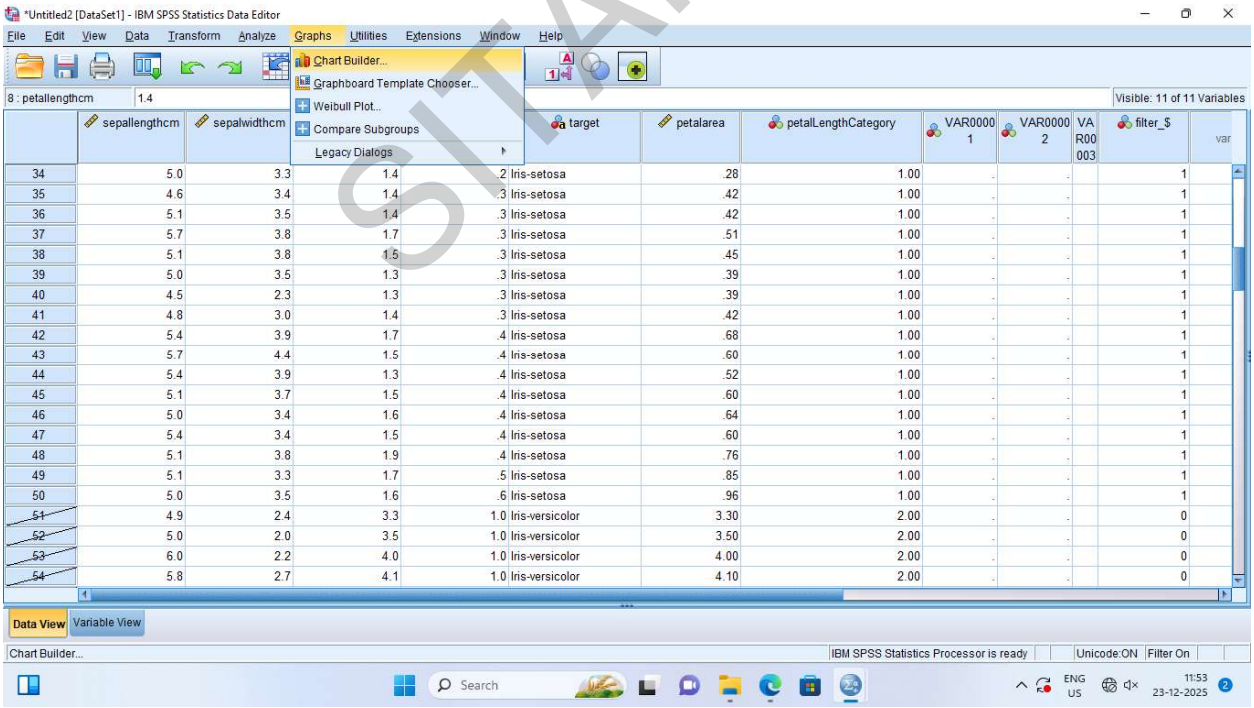

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
petalLengthCategory.
VARIABLE LABELS petalLengthCategory 'petal length category(cm)'.
EXECUTE.
SORT CASES BY target.
SPLIT FILE SEPARATE BY target.
SPLIT FILE OFF.
SPLIT FILE OFF.
USE ALL.
COMPUTE filter_$=(target = "Iris-Setosa").
VARIABLE LABELS filter_$ 'target = "Iris-Setosa" (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
FILTER OFF.
USE ALL.
EXECUTE.
SPLIT FILE OFF.
USE ALL.
IBM SPSS Statistics Processor is ready
Unicode:ON
Double-click to activate
sepallengthcm
filter_$
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:51
23-12-2025

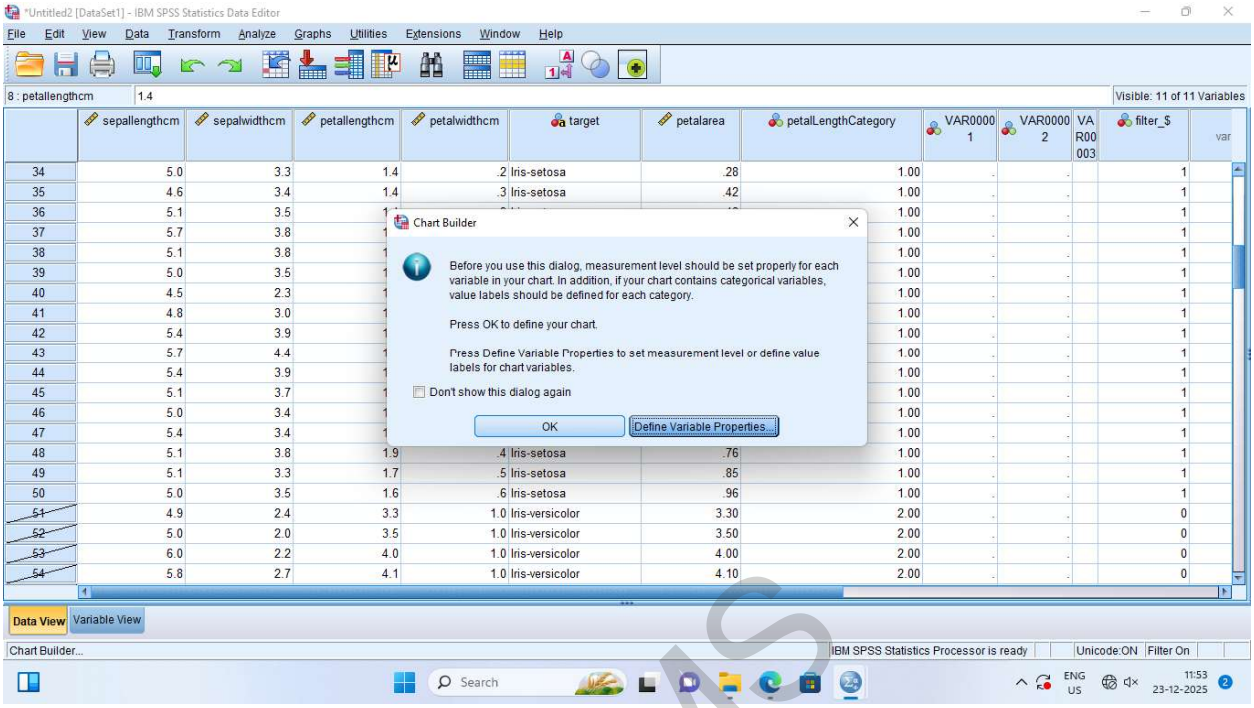

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Reports
Descriptive Statistics
Bayesian Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Direct Marketing
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
sepallengthcm
petalarea
petalLengthCategory
VAR00001
VAR00002
filter_$
Data View
Variable View
Frequencies...
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:53
23-12-2025

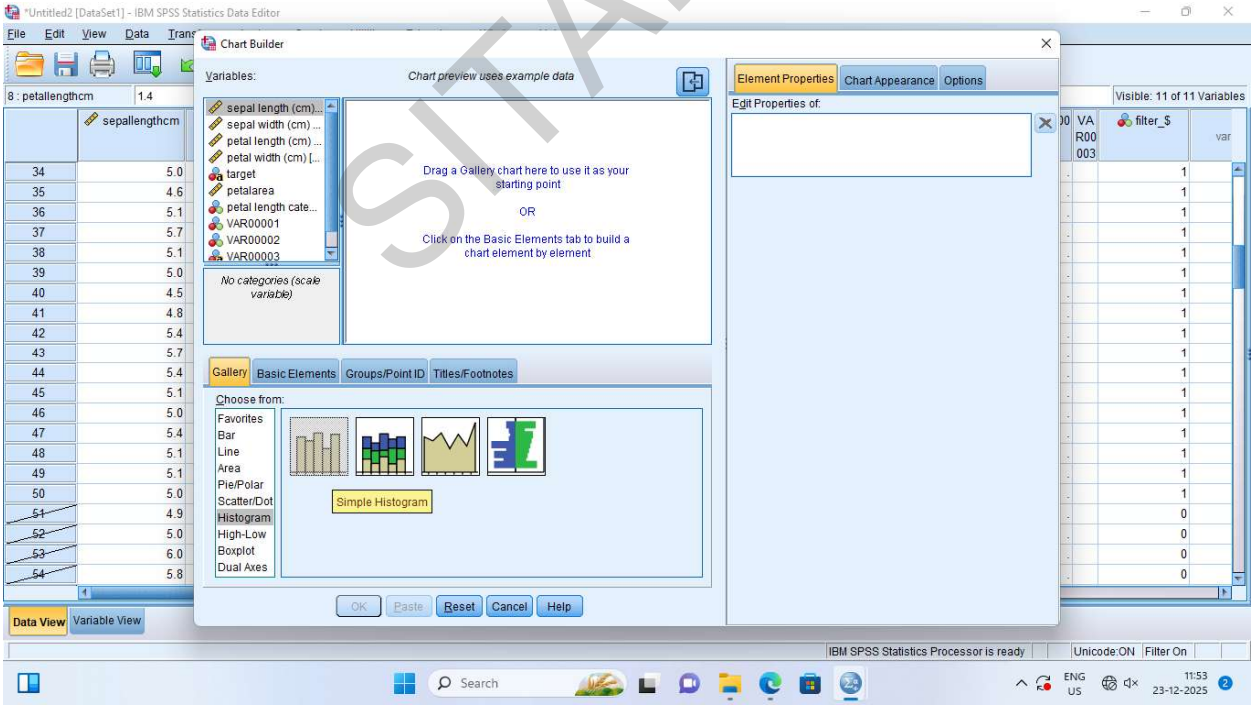

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory VAR00001 VAR00002 VAR00003 filter_$ var
Frequencies
Variable(s):
target
sepal length (cm)...
sepal width (cm) ...
petal length (cm) ...
petal width (cm) [...
petalarea
petal length cate...
VAR00001
VAR00002
VAR00003
Statistics...
Charts...
Format...
Style...
Bootstrap...
Display frequency tables
OK Paste Reset Cancel Help
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:53 23-12-2025

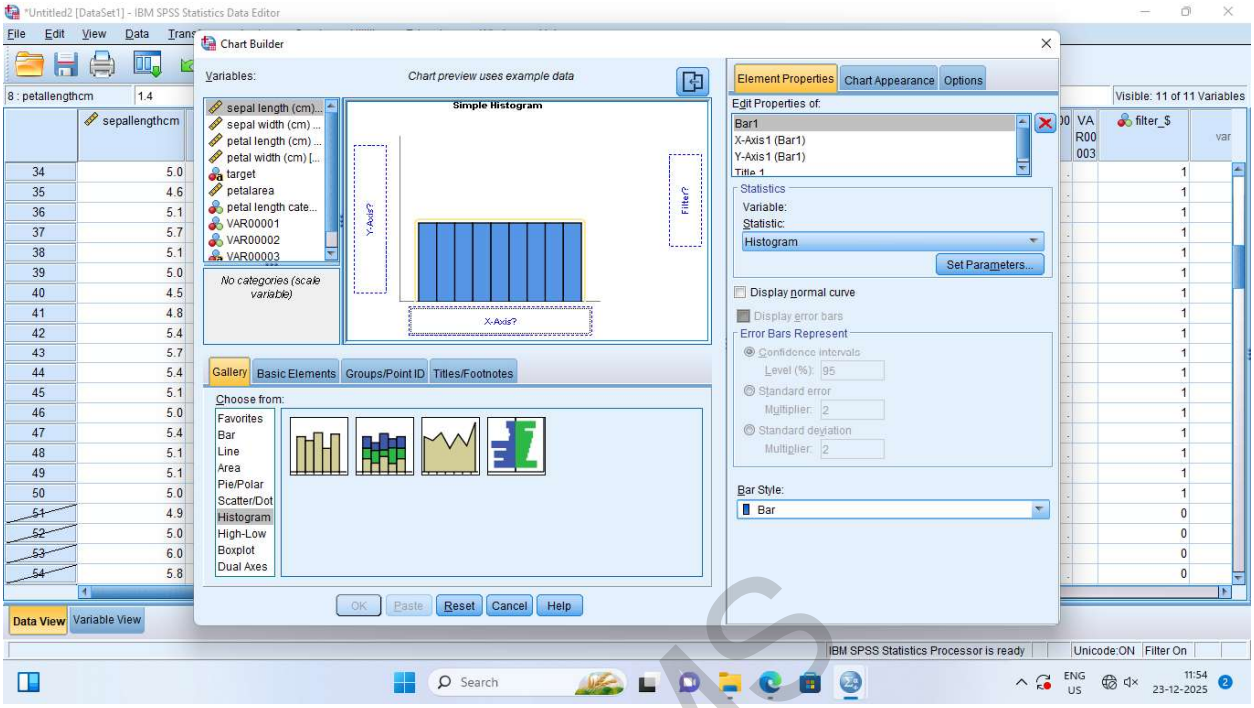

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
target
Frequencies
[DataSet1]
Statistics
target
N Valid 50
Missing 0
target
Frequency Percent Valid Percent Cumulative Percent
Valid Iris-setosa 50 100.0 100.0 100.0
IBM SPSS Statistics Processor is ready
Unicode:ON
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:53 23-12-2025

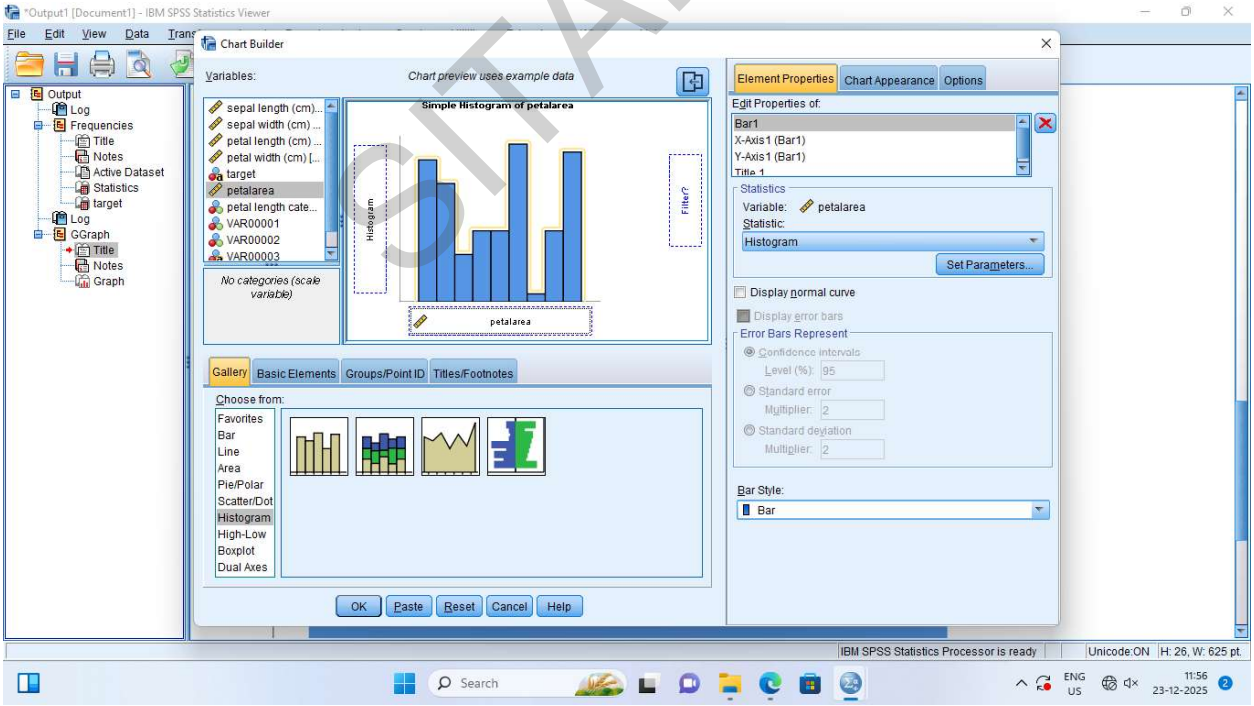

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory VAR00001 VAR00002 VAR00003 filter_$ var
34 5.0 3.3 1.4 .2 Iris-setosa .28 1.00 1
35 4.6 3.4 1.4 .3 Iris-setosa .42 1.00 1
36 5.1 3.5 1.4 .3 Iris-setosa .42 1.00 1
37 5.7 3.8 1.7 .3 Iris-setosa .51 1.00 1
38 5.1 3.8 1.5 .3 Iris-setosa .45 1.00 1
39 5.0 3.5 1.3 .3 Iris-setosa .39 1.00 1
40 4.5 2.3 1.3 .3 Iris-setosa .39 1.00 1
41 4.8 3.0 1.4 .3 Iris-setosa .42 1.00 1
42 5.4 3.9 1.7 .4 Iris-setosa .68 1.00 1
43 5.7 4.4 1.5 .4 Iris-setosa .60 1.00 1
44 5.4 3.9 1.3 .4 Iris-setosa .52 1.00 1
45 5.1 3.7 1.5 .4 Iris-setosa .60 1.00 1
46 5.0 3.4 1.6 .4 Iris-setosa .64 1.00 1
47 5.4 3.4 1.5 .4 Iris-setosa .60 1.00 1
48 5.1 3.8 1.9 .4 Iris-setosa .76 1.00 1
49 5.1 3.3 1.7 .5 Iris-setosa .85 1.00 1
50 5.0 3.5 1.6 .6 Iris-setosa .96 1.00 1
51 4.9 2.4 3.3 1.0 Iris-versicolor 3.30 2.00 0
52 5.0 2.0 3.5 1.0 Iris-versicolor 3.50 2.00 0
53 6.0 2.2 4.0 1.0 Iris-versicolor 4.00 2.00 0
54 5.8 2.7 4.1 1.0 Iris-versicolor 4.10 2.00 0
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:53 23-12-2025

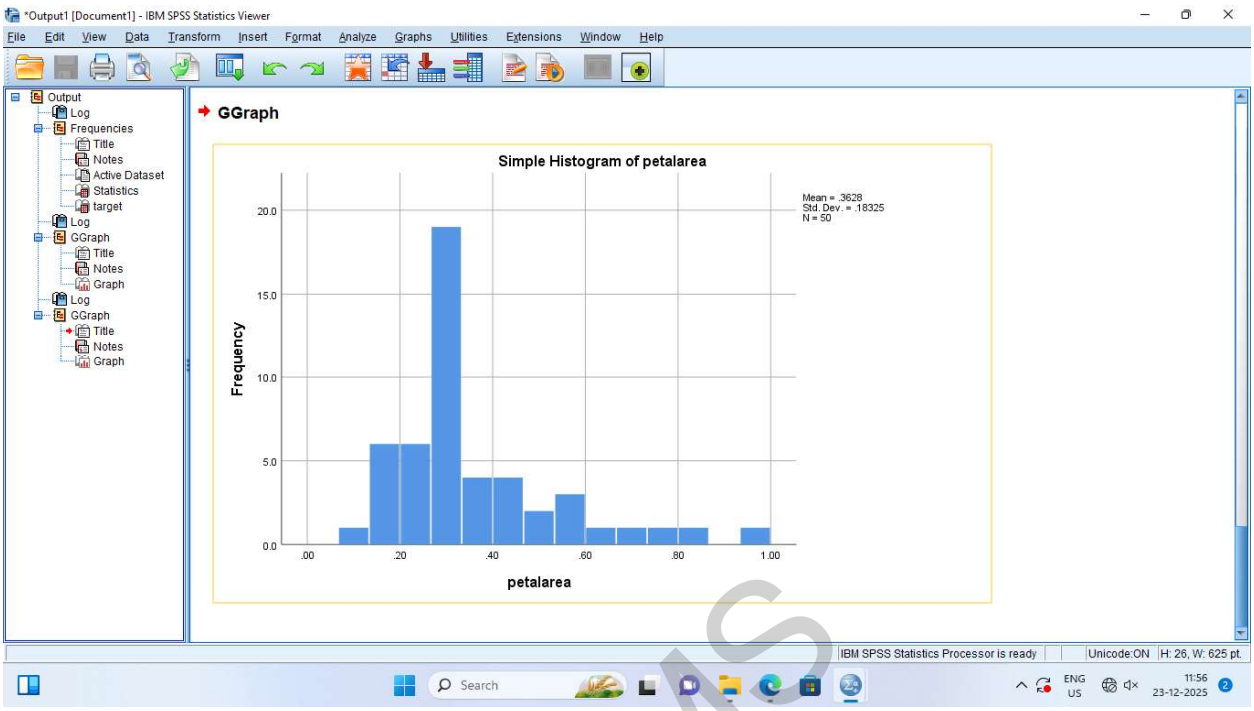

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Chart Builder...
Graphboard Template Chooser...
Weibull Plot...
Compare Subgroups
Legacy Dialogs
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
sepallengthcm sepalwidthcm target petalarea petalLengthCategory VAR00001 VAR00002 VAR00003 filter_$ var
34 5.0 3.3 1.4 .2 Iris-setosa .28 1.00 1
35 4.6 3.4 1.4 .3 Iris-setosa .42 1.00 1
36 5.1 3.5 1.4 .3 Iris-setosa .42 1.00 1
37 5.7 3.8 1.7 .3 Iris-setosa .51 1.00 1
38 5.1 3.8 1.5 .3 Iris-setosa .45 1.00 1
39 5.0 3.5 1.3 .3 Iris-setosa .39 1.00 1
40 4.5 2.3 1.3 .3 Iris-setosa .39 1.00 1
41 4.8 3.0 1.4 .3 Iris-setosa .42 1.00 1
42 5.4 3.9 1.7 .4 Iris-setosa .68 1.00 1
43 5.7 4.4 1.5 .4 Iris-setosa .60 1.00 1
44 5.4 3.9 1.3 .4 Iris-setosa .52 1.00 1
45 5.1 3.7 1.5 .4 Iris-setosa .60 1.00 1
46 5.0 3.4 1.6 .4 Iris-setosa .64 1.00 1
47 5.4 3.4 1.5 .4 Iris-setosa .60 1.00 1
48 5.1 3.8 1.9 .4 Iris-setosa .76 1.00 1
49 5.1 3.3 1.7 .5 Iris-setosa .85 1.00 1
50 5.0 3.5 1.6 .6 Iris-setosa .96 1.00 1
51 4.9 2.4 3.3 1.0 Iris-versicolor 3.30 2.00 0
52 5.0 2.0 3.5 1.0 Iris-versicolor 3.50 2.00 0
53 6.0 2.2 4.0 1.0 Iris-versicolor 4.00 2.00 0
54 5.8 2.7 4.1 1.0 Iris-versicolor 4.10 2.00 0
Data View Variable View
Chart Builder...
IBM SPSS Statistics Processor is ready
Unicode:ON Filter On
Search
ENG US
11:53 23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
8 : petallengthcm 1.4
Visible: 11 of 11 Variables
sepallengthcm sepalwidthcm petallengthcm petalwidthcm target petalarea petalLengthCategory VAR00001 VAR00002 VAR00003 filter_$ var
Chart Builder
Before you use this dialog, measurement level should be set properly for each variable in your chart. In addition, if your chart contains categorical variables, value labels should be defined for each category.
Press OK to define your chart.
Press Define Variable Properties to set measurement level or define value labels for chart variables.
Don't show this dialog again
OK
Define Variable Properties...
Data View
Variable View
Chart Builder...
IBM SPSS Statistics Processor is ready
Unicode:ON
Filter On
Search
ENG US
11:53 23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
Chart Builder
Variables:
Chart preview uses example data
sepal length (cm)...
sepal width (cm) ...
petal length (cm) ...
petal width (cm) [...
target
petalarea
petal length cate...
VAR00001
VAR00002
VAR00003
No categories (scale variable)
Drag a Gallery chart here to use it as your starting point
OR
Click on the Basic Elements tab to build a chart element by element
Element Properties
Chart Appearance
Options
Edit Properties of:
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
Simple Histogram
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Filter On
Search
ENG US
11:53 23-12-2025


*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data
Chart Builder
Variables:
Chart preview uses example data
Simple Histogram
sepal length (cm)...
sepal width (cm) ...
petal length (cm) ...
petal width (cm) [...
target
petalarea
petal length cate...
VAR00001
VAR00002
VAR00003
No categories (scale variable)
Y-Axis?
X-Axis?
Filter?
Element Properties
Chart Appearance
Options
Edit Properties of:
Bar1
X-Axis1 (Bar1)
Y-Axis1 (Bar1)
Statistics
Variable:
Statistic:
Histogram
Set Parameters...
Display normal curve
Display error bars
Error Bars Represent
Confidence intervals
Level (%): 95
Standard error
Multiplier: 2
Standard deviation
Multiplier: 2
Bar Style:
Bar
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Filter On
Search
ENG
US
11:54
23-12-2025


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
target
Log
GGraph
Title
Notes
Graph
Chart Builder
Variables:
Chart preview uses example data
Simple Histogram of petalarea
sepal length (cm)...
sepal width (cm) ...
petal length (cm) ...
petal width (cm) [...
target
petalarea
petal length cate...
VAR00001
VAR00002
VAR00003
No categories (scale variable)
Histogram
petalarea
Filter?
Element Properties
Chart Appearance
Options
Edit Properties of:
Bar1
X-Axis1 (Bar1)
Y-Axis1 (Bar1)
Statistics
Variable: petalarea
Statistic:
Histogram
Set Parameters...
Display normal curve
Display error bars
Error Bars Represent
Confidence intervals
Level (%): 95
Standard error
Multiplier: 2
Standard deviation
Multiplier: 2
Bar Style:
Bar
Gallery
Basic Elements
Groups/Point ID
Titles/Footnotes
Choose from:
Favorites
Bar
Line
Area
Pie/Polar
Scatter/Dot
Histogram
High-Low
Boxplot
Dual Axes
OK
Paste
Reset
Cancel
Help
IBM SPSS Statistics Processor is ready
Unicode:ON
H: 26, W: 625 pt.
Search
ENG
US
11:56
23-12-2025


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help
Output
Log
Frequencies
Title
Notes
Active Dataset
Statistics
target
Log
GGraph
Title
Notes
Graph
Log
GGraph
Title
Notes
Graph
GGraph
Simple Histogram of petalarea
Mean = .3628
Std. Dev. = .18325
N = 50
Frequency
20.0
15.0
10.0
5.0
0.0
.00
.20
.40
.60
.80
1.00
petalarea
IBM SPSS Statistics Processor is ready
Unicode:ON
H: 26, W: 625 pt.
Search
ENG
US
11:56
23-12-2025

Results

1.A: The sorting operation was successfully performed on the dataset.

1.B: The "Split File" command was successfully applied.

1.C: The computation of a new variable was completed successfully.

1.D: The recoding operation was successfully executed using the "Recode into Different Variables" method.

1.E: The selection of cases was successfully performed based on the specific condition provided.

EXP2: Chi- Square Test (Parametric and Non-Parametric)

Aim

To apply the **Chi-Square (χ^2) Test** in SPSS to:

1. Test whether the observed frequency distribution of a single categorical variable significantly differs from a hypothesized (expected) distribution (**Goodness-of-Fit Test-non-parametric**).
2. Determine whether there is a significant association between two categorical variables (**Test of Independence-parametric**).

Description:

The Chi-Square (χ^2) Test is a non-parametric statistical method used to analyze categorical data. It compares observed frequencies with expected frequencies to determine if significant differences exist (Goodness-of-Fit) or if two variables are associated (Test of Independence).

Types

1. **Chi-Square Goodness-of-Fit Test**
   - Used for **one categorical variable**.
   - Tests if the observed distribution matches a specified (often uniform) expected distribution.
   - Example: Is a die fair? (expected: equal counts for 1–6).
2. **Chi-Square Test of Independence**
   - Used for **two categorical variables** in a contingency table.
   - Tests if the variables are independent or associated.
   - Example: Is there an association between gender and preferred sport?

Formula

$$\chi^2 = \Sigma \left[(O_i - E_i)^2 / E_i\right]$$

Where:

- O_i = Observed frequency
- E_i = Expected frequency
- Summation over all categories/cells

Procedure:

1. Chi-Square Goodness-of-Fit Test

Step by step procedure:

1. Prepare data: One categorical variable (preferably **numeric** codes, e.g., 1,2,3... with value labels). Example: Dice_Roll (1 to 6).
2. Go to **Analyze > Nonparametric Tests > Legacy Dialogs > Chi-Square**.
3. Move the categorical variable (e.g., Dice_Roll) to **Test Variable List**.
4. Under **Expected Values**:
    - Select **All categories equal** (for uniform distribution, e.g., fair die).
    - OR select **Values** and enter custom expected proportions (they must sum to 1 if proportions, or counts).
5. Click **OK**.

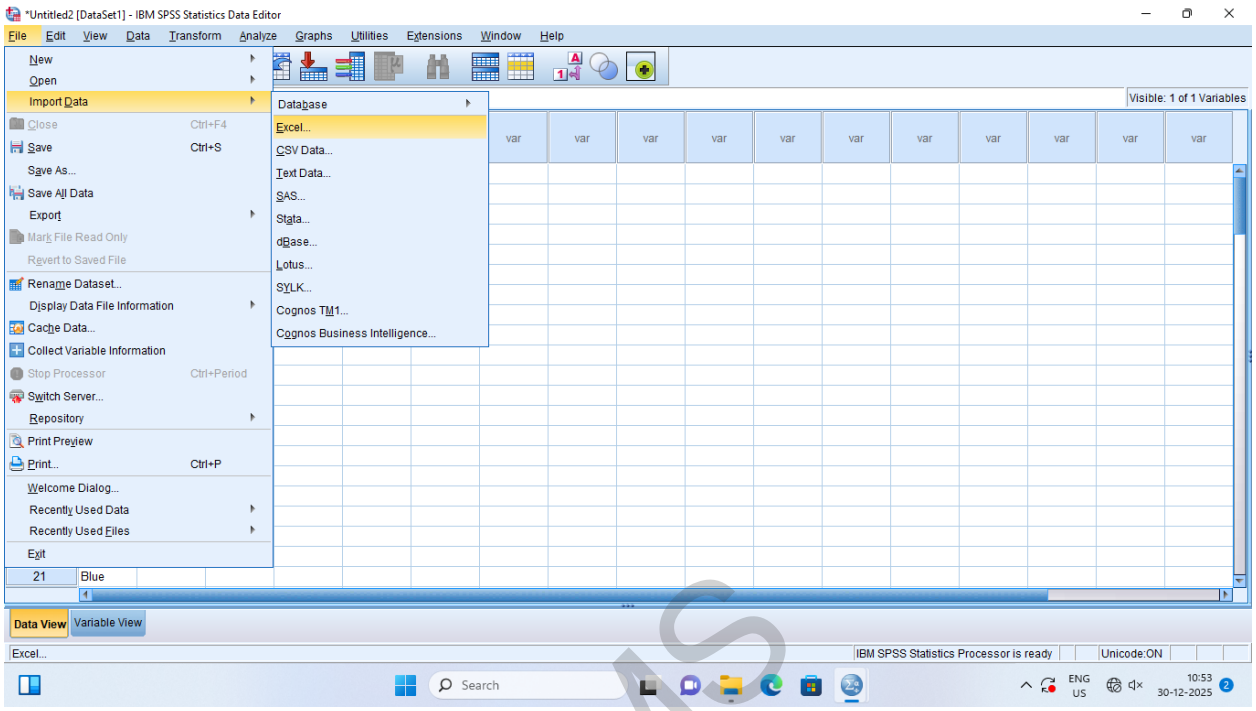
Interpretation:

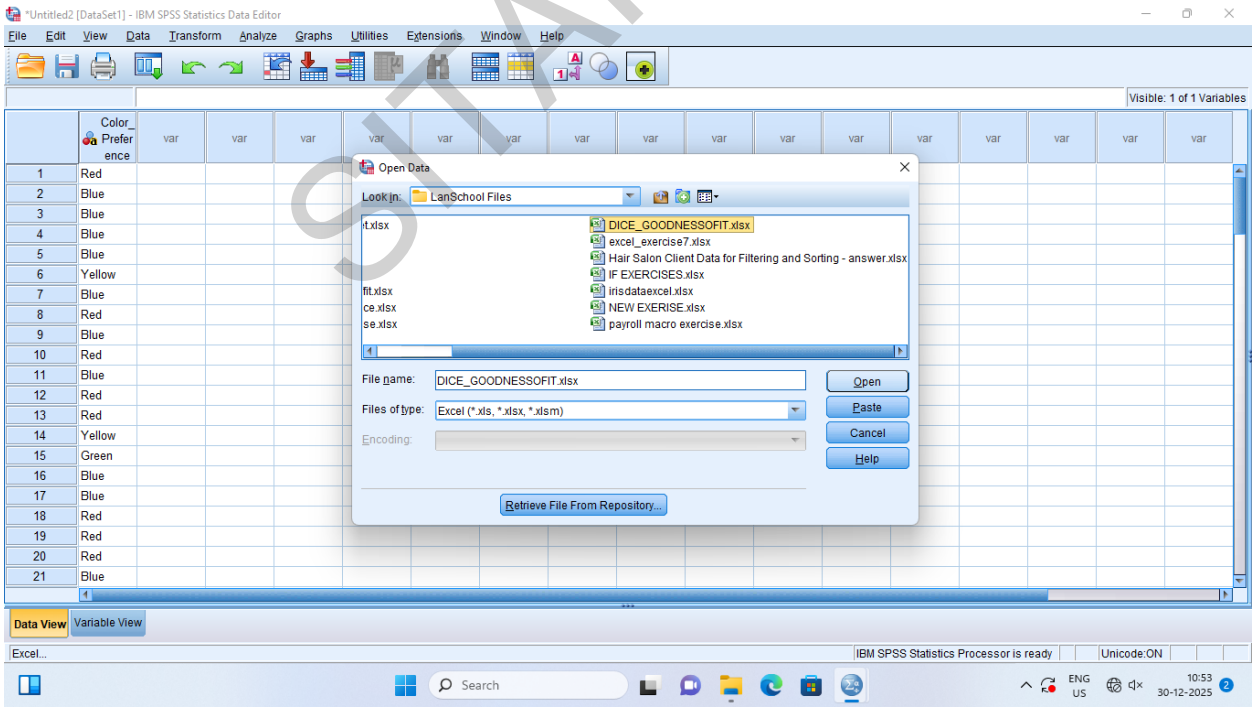
- Compares observed vs. expected frequencies.
- p-value (Asymp. Sig.): If $p < 0.05$, distribution differs significantly from hypothesized.

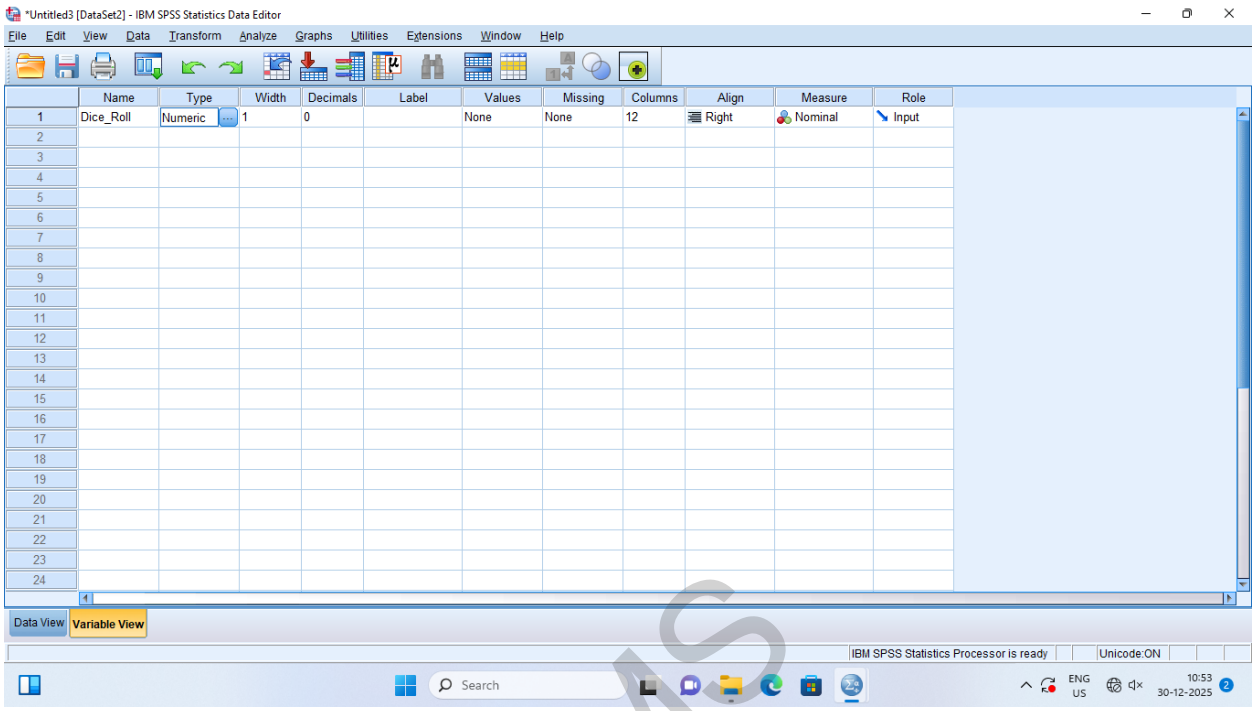
p-value	Decision	Meaning
$p \leq 0.05$	Reject null hypothesis	Significant → Die is likely unfair
$p > 0.05$	Fail to reject null hypothesis	Not significant → Die appears fair

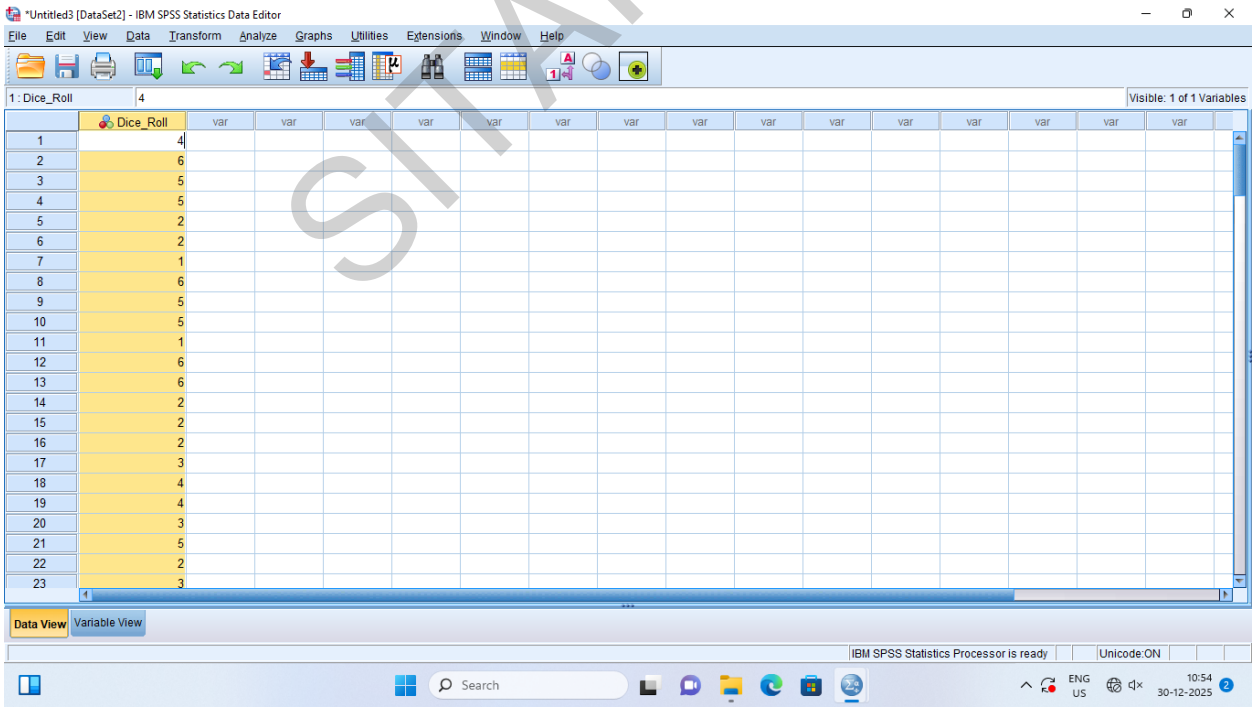
Expected Output:

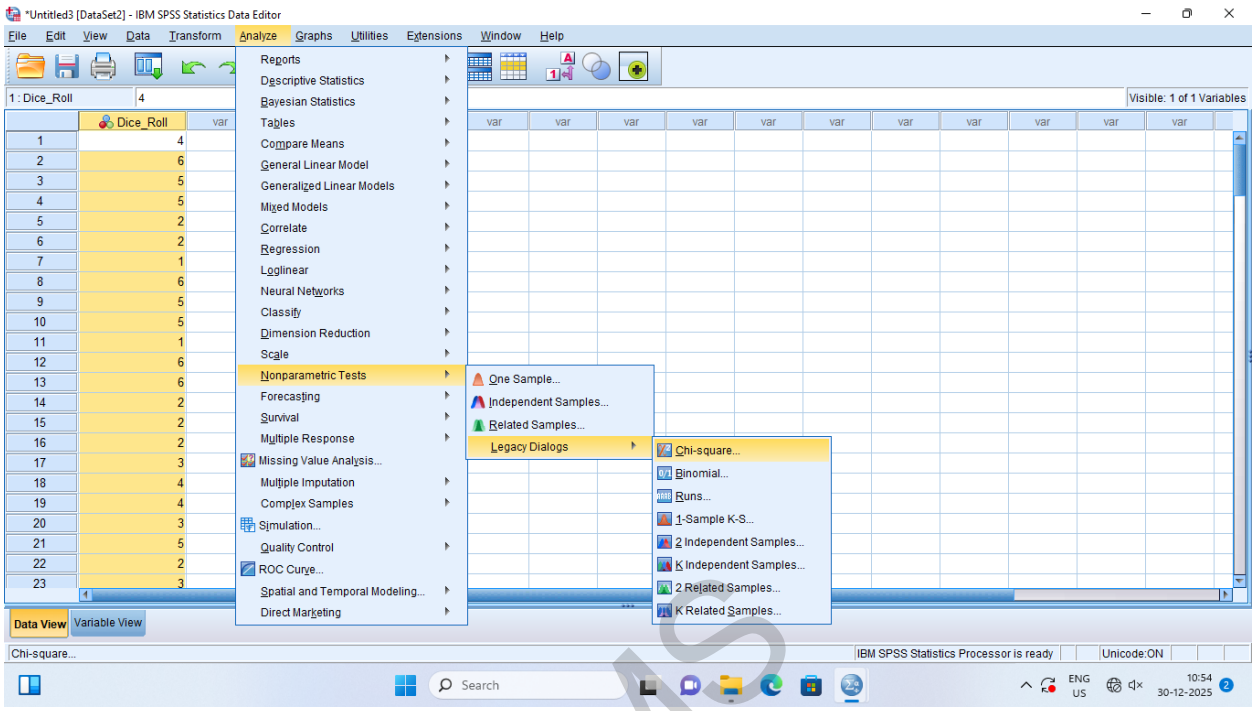
- Observed N and Expected N for each category.
- Chi-Square value, df, Asymp. Sig. (p-value).
- Footnote about expected counts <5 (if any).

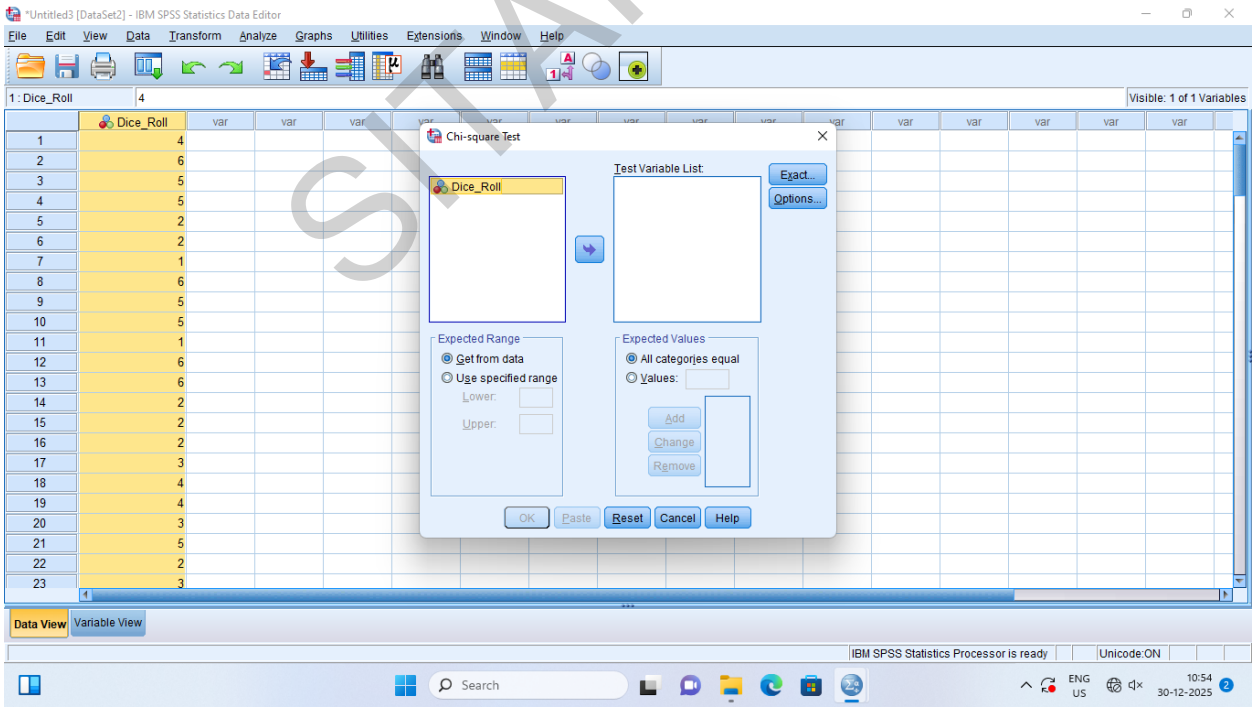

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
New
Open
Import Data
Close Ctrl+F4
Save Ctrl+S
Save As...
Save All Data
Export
Mark File Read Only
Revert to Saved File
Rename Dataset...
Display Data File Information
Cache Data...
Collect Variable Information
Stop Processor Ctrl+Period
Switch Server...
Repository
Print Preview
Print... Ctrl+P
Welcome Dialog...
Recently Used Data
Recently Used Files
Exit
Database
Excel...
CSV Data...
Text Data...
SAS...
Stata...
dBase...
Lotus...
SYLK...
Cognos TM1...
Cognos Business Intelligence...
Visible: 1 of 1 Variables
var
21 Blue
Data View Variable View
Excel...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:53 30-12-2025

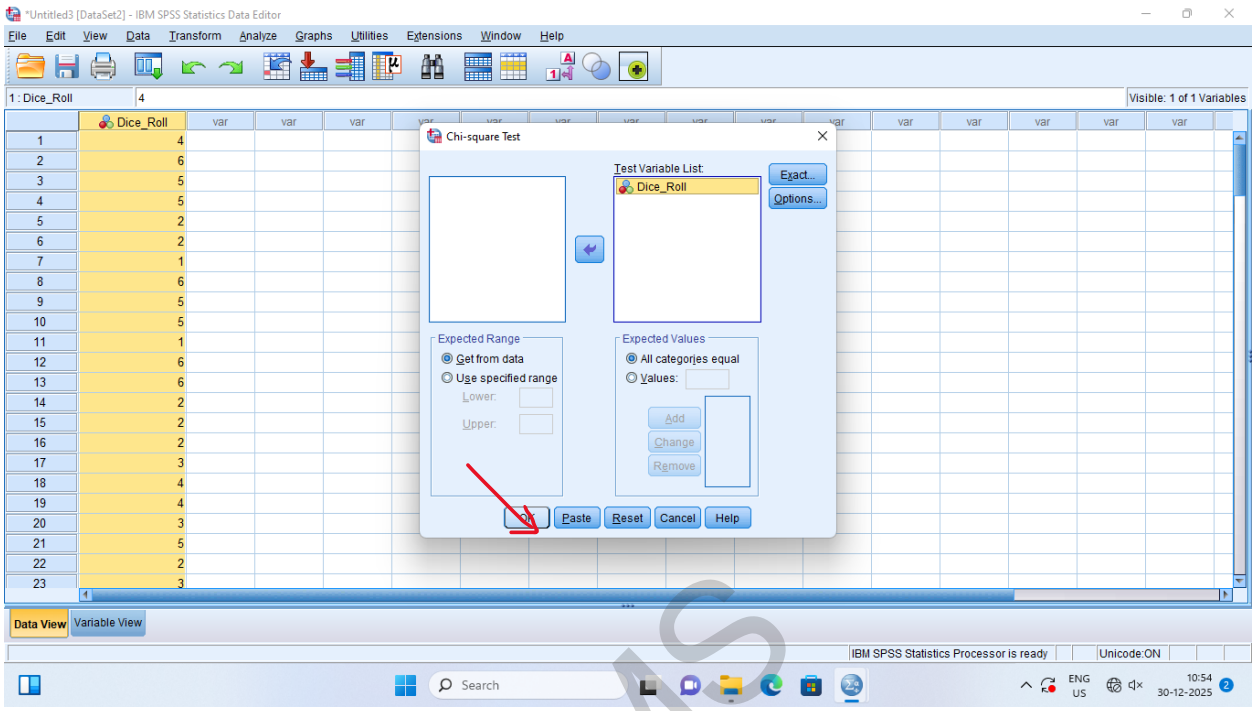

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 1 of 1 Variables
Color_ Preference
1 Red
2 Blue
3 Blue
4 Blue
5 Blue
6 Yellow
7 Blue
8 Red
9 Blue
10 Red
11 Blue
12 Red
13 Red
14 Yellow
15 Green
16 Blue
17 Blue
18 Red
19 Red
20 Red
21 Blue
Open Data
Look in: LanSchool Files
DICE_GOODNESSOFIT.xlsx
excel_exercise7.xlsx
Hair Salon Client Data for Filtering and Sorting - answer.xlsx
IF EXERCISES.xlsx
irisdataexcel.xlsx
NEW EXERISE.xlsx
payroll macro exercise.xlsx
File name: DICE_GOODNESSOFIT.xlsx
Files of type: Excel (*.xls, *.xlsx, *.xlsm)
Encoding:
Open
Paste
Cancel
Help
Retrieve File From Repository...
Data View Variable View
Excel...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:53 30-12-2025


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
1 Dice_Roll Numeric 1 0 None None 12 Right Nominal Input
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:53 30-12-2025


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
1 : Dice_Roll
4
Visible: 1 of 1 Variables
Dice_Roll
1 4
2 6
3 5
4 5
5 2
6 2
7 1
8 6
9 5
10 5
11 1
12 6
13 6
14 2
15 2
16 2
17 3
18 4
19 4
20 3
21 5
22 2
23 3
Data View Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:54 30-12-2025


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Reports
Descriptive Statistics
Bayesian Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Direct Marketing
One Sample...
Independent Samples...
Related Samples...
Legacy Dialogs
Chi-square...
Binomial...
Runs...
1-Sample K-S...
2 Independent Samples...
K Independent Samples...
2 Related Samples...
K Related Samples...
Dice_Roll
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
Chi-square Test
Test Variable List:
Dice_Roll
Exact...
Options...
Expected Range
Get from data
Use specified range
Lower:
Upper:
Expected Values
All categories equal
Values:
Add
Change
Remove
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON


*Untitled3 [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
1 : Dice_Roll
4
Visible: 1 of 1 Variables
Dice_Roll
Chi-square Test
Test Variable List:
Dice_Roll
Exact...
Options...
Expected Range
Get from data
Use specified range
Lower:
Upper:
Expected Values
All categories equal
Values:
Add
Change
Remove
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:54
30-12-2025

Output:

```
GET DATA
  /TYPE=XLSX
  /FILE='D:\palab\c_goodnessoffit.xlsx'
  /SHEET=name 'c_goodnessoffit'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet2 WINDOW=FRONT.
NPAR TESTS
  /CHISQUARE=Color_Preference
  /EXPECTED=EQUAL
  /MISSING ANALYSIS.


NPAR TESTS
  /CHISQUARE=Dice_Roll
  /EXPECTED=EQUAL
  /MISSING ANALYSIS.
```


NPar Tests

Chi-Square Test

Frequencies

Dice_Roll

	Observed N	Expected N	Residual
1	18	25.0	-7.0
2	20	25.0	-5.0
3	29	25.0	4.0
4	24	25.0	-1.0
5	28	25.0	3.0
6	31	25.0	6.0
Total	150		

Test Statistics	
	Dice_Roll
Chi-Square	5.440[a]
df	5
Asymp. Sig.	.365
a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.0.	

2. Chi-Square Test of Independence

Step by step procedure:

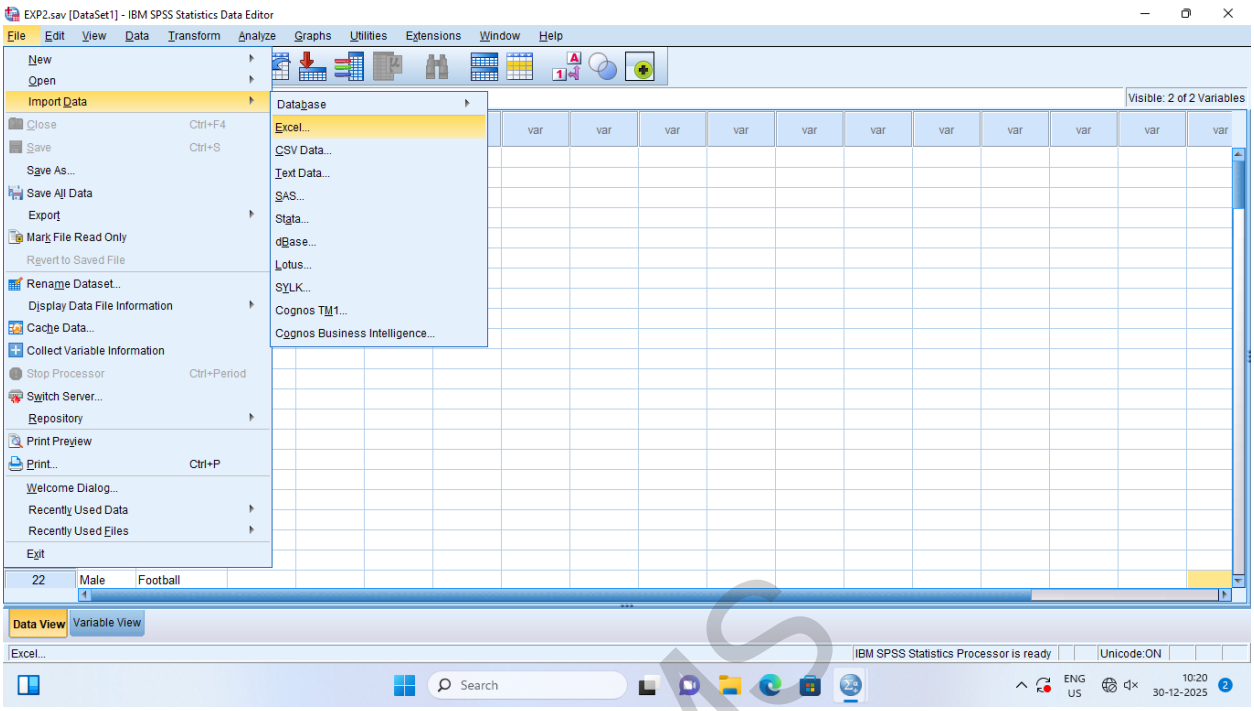
1. Go to **Analyze > Descriptive Statistics > Crosstabs**.
2. Move one variable to **Row(s)** and the other to **Column(s)**.
3. Click **Statistics** button:
    - Check **Chi-square**.
    - Optionally check **Phi and Cramer's V** (effect size).
    - Click **Continue**.
4. Click **Cells** button:
    - Check **Observed** and **Expected** counts.
    - Check **Row** or **Column** percentages (for interpretation).
    - Click **Continue**.
5. Click **OK**.

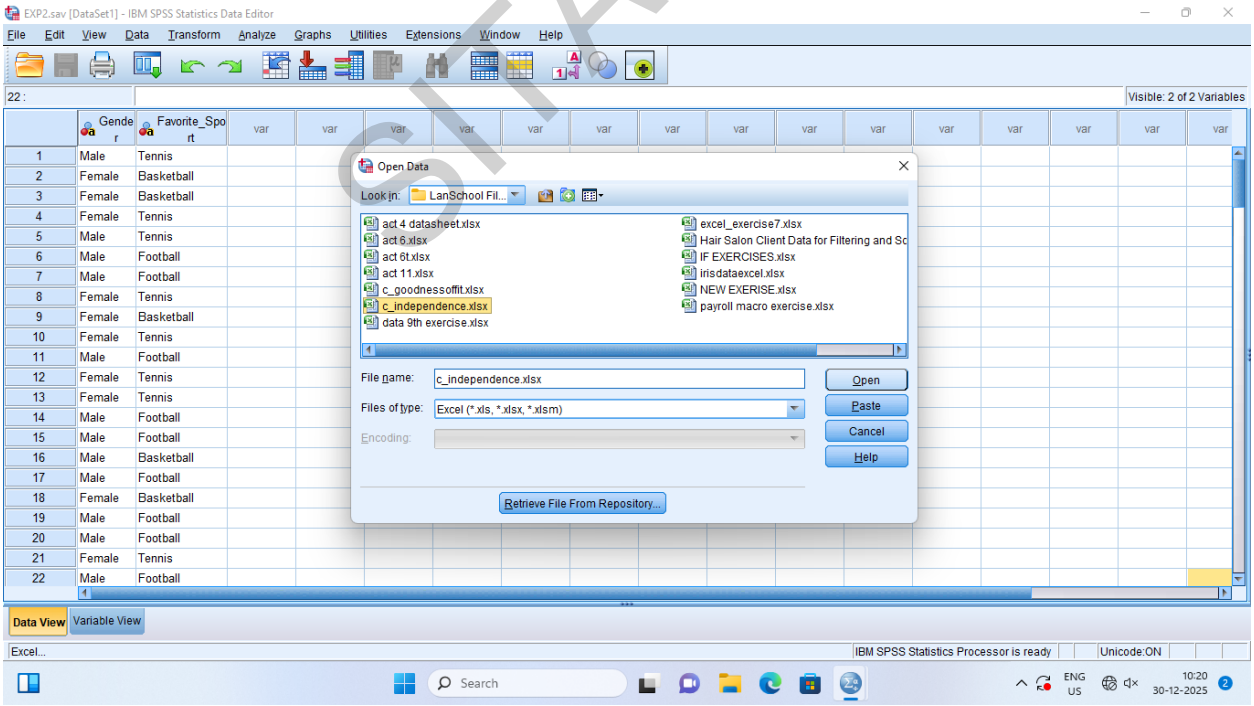
Interpretation:

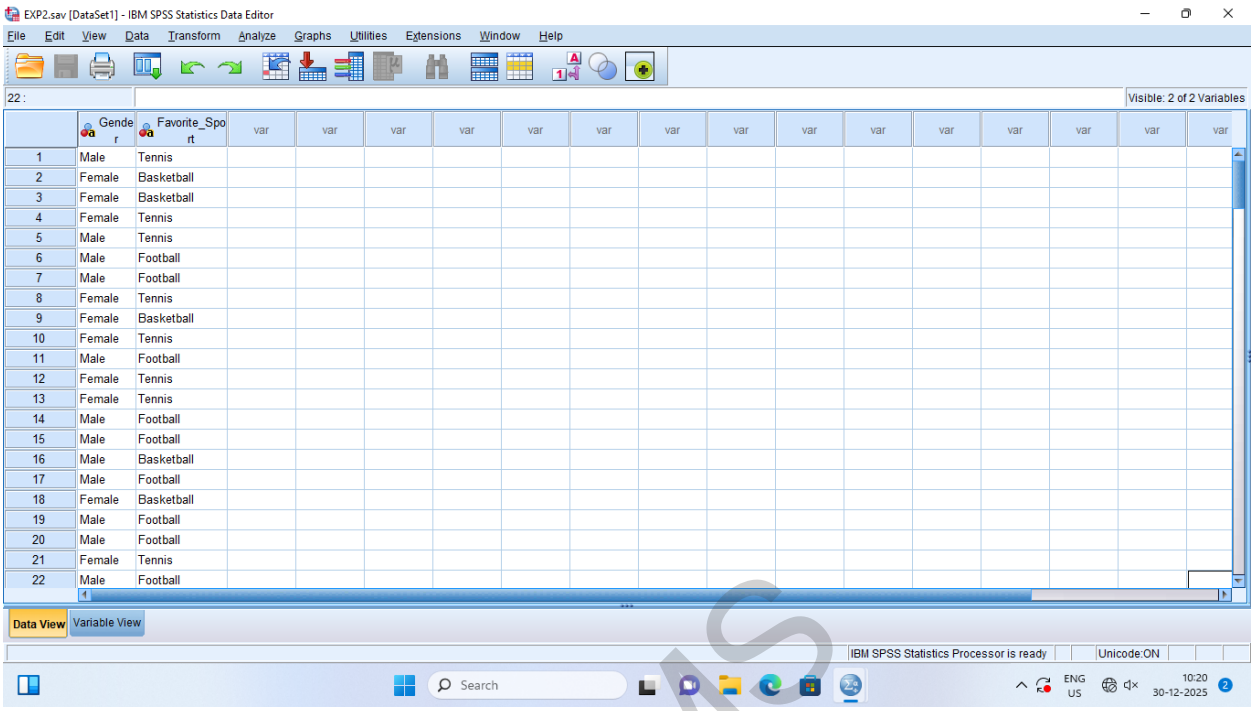
- Look at the **Crosstabulation table**: Compare observed vs. expected counts.
- In **Chi-Square Tests** table: Focus on **Pearson Chi-Square** row.
    - **Asymp. Sig. (2-sided)** = p-value.
    - If $p < 0.05$, significant association (variables are related).
    - Check footnotes for warnings (e.g., if expected counts < 5 in >20% of cells, results may be unreliable).
- Effect size: Phi or Cramer's V (0.1 small, 0.3 medium, 0.5 large).

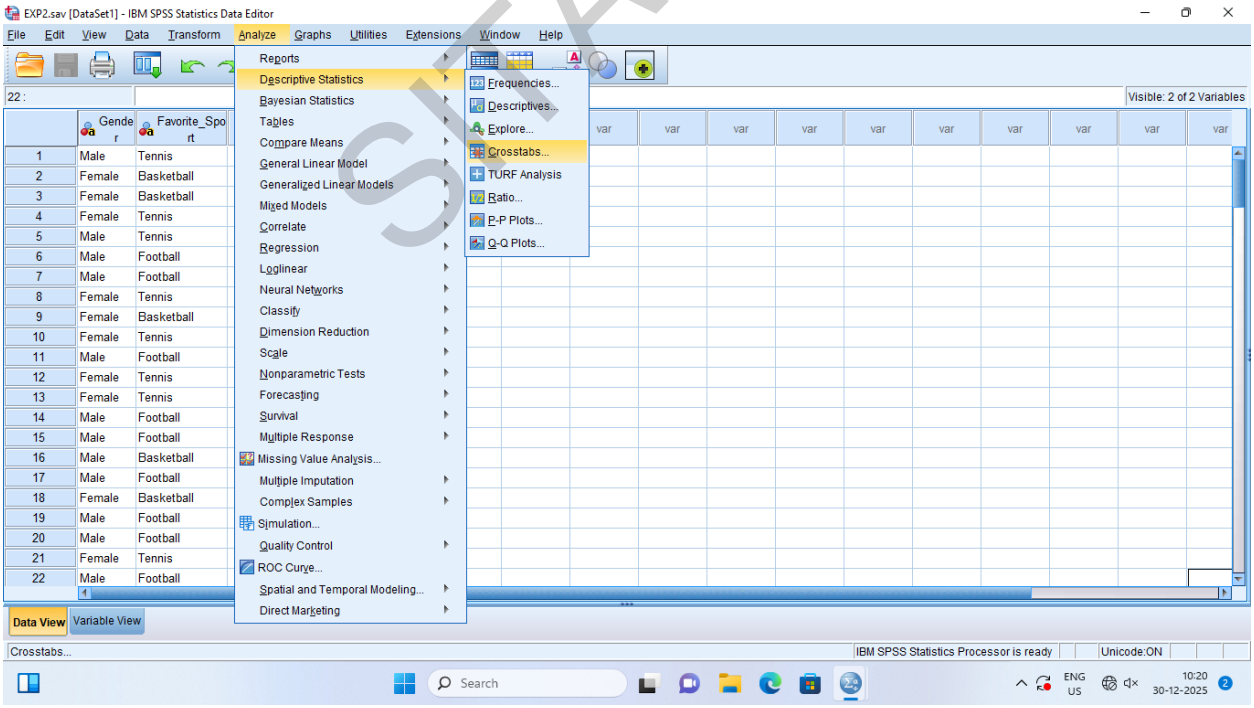
Expected Output:

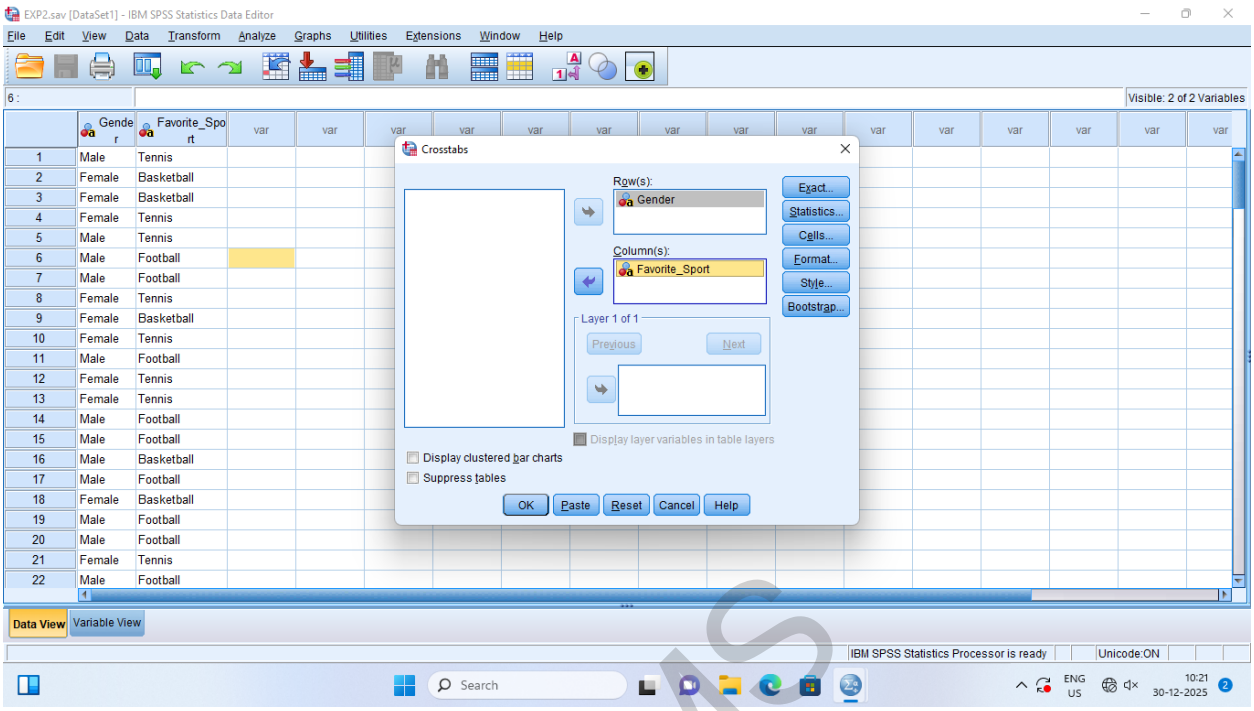
- **Crosstabulation table**: Observed vs. Expected counts and percentages.
- **Chi-Square Tests table**: Pearson Chi-Square row → Value, df, Asymp. Sig. (2-sided) p-value.
- Effect size (if selected): Cramer's V (0.1 = small, 0.3 = medium, 0.5 = large).
- Footnotes: Warn if expected counts are low.

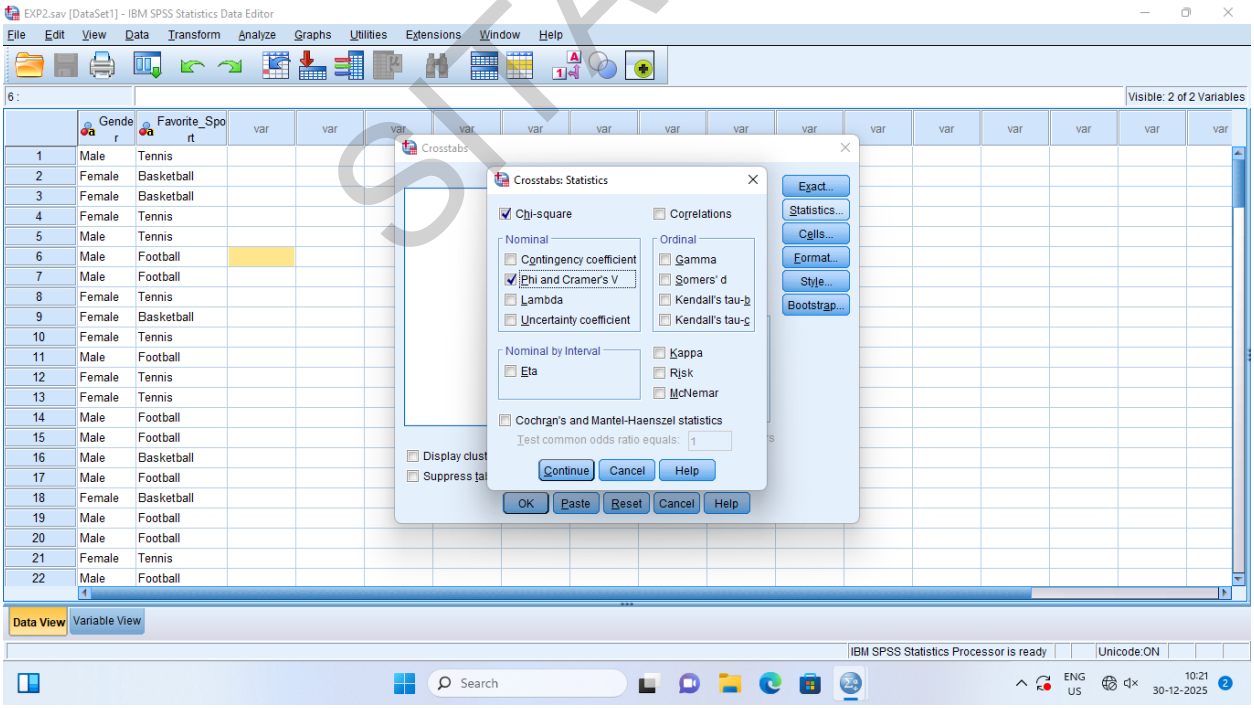

EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
New
Open
Import Data
Close Ctrl+F4
Save Ctrl+S
Save As...
Save All Data
Export
Mark File Read Only
Revert to Saved File
Rename Dataset...
Display Data File Information
Cache Data...
Collect Variable Information
Stop Processor Ctrl+Period
Switch Server...
Repository
Print Preview
Print... Ctrl+P
Welcome Dialog...
Recently Used Data
Recently Used Files
Exit
Database
Excel...
CSV Data...
Text Data...
SAS...
Stata...
dBase...
Lotus...
SYLK...
Cognos TM1...
Cognos Business Intelligence...
Visible: 2 of 2 Variables
22 Male Football
Data View Variable View
Excel...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:20 30-12-2025

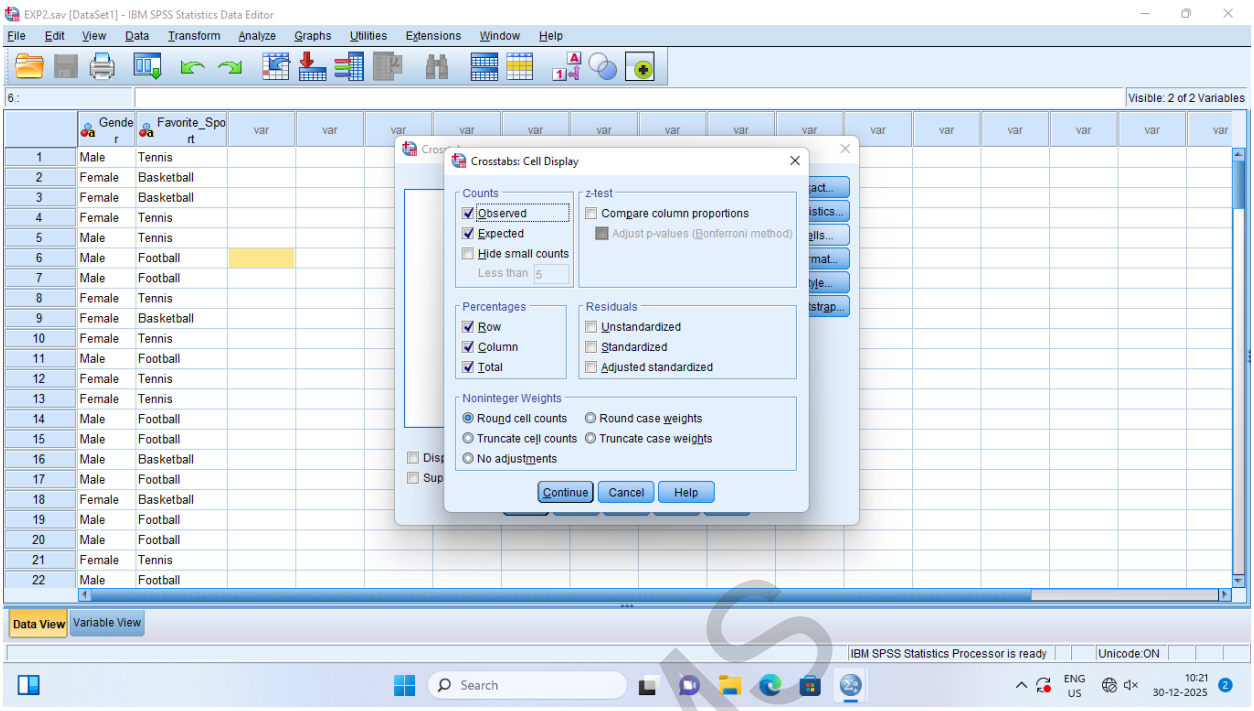

EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
22 :
Visible: 2 of 2 Variables
Gender Favorite_Sport
1 Male Tennis
2 Female Basketball
3 Female Basketball
4 Female Tennis
5 Male Tennis
6 Male Football
7 Male Football
8 Female Tennis
9 Female Basketball
10 Female Tennis
11 Male Football
12 Female Tennis
13 Female Tennis
14 Male Football
15 Male Football
16 Male Basketball
17 Male Football
18 Female Basketball
19 Male Football
20 Male Football
21 Female Tennis
22 Male Football
Open Data
Look in: LanSchool Fil...
act 4 datasheet.xlsx
act 6.xlsx
act 6t.xlsx
act 11.xlsx
c_goodnessoffit.xlsx
c_independence.xlsx
data 9th exercise.xlsx
excel_exercise7.xlsx
Hair Salon Client Data for Filtering and Sc
IF EXERCISES.xlsx
irisdataexcel.xlsx
NEW EXERISE.xlsx
payroll macro exercise.xlsx
File name: c_independence.xlsx
Files of type: Excel (*.xls, *.xlsx, *.xlsm)
Encoding:
Open
Paste
Cancel
Help
Retrieve File From Repository...
Data View Variable View
Excel...
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:20 30-12-2025

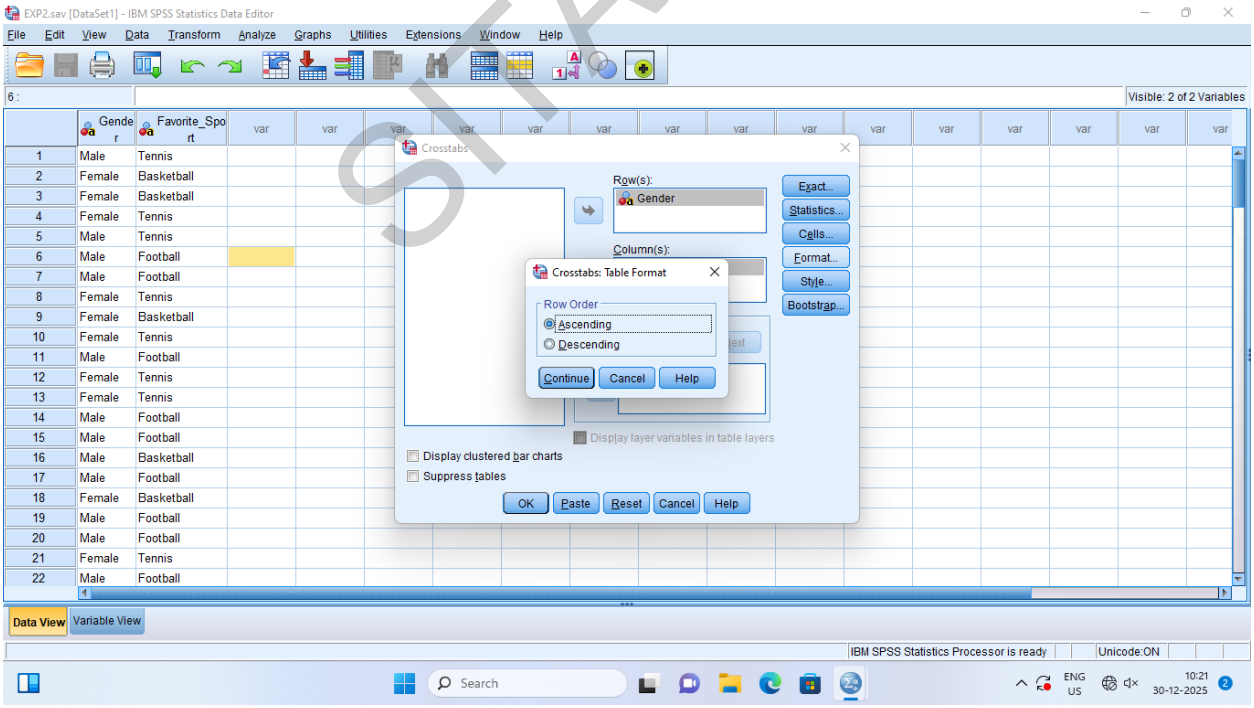

EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
22 :
Visible: 2 of 2 Variables
Gender
Favorite_Sport
var
1 Male Tennis
2 Female Basketball
3 Female Basketball
4 Female Tennis
5 Male Tennis
6 Male Football
7 Male Football
8 Female Tennis
9 Female Basketball
10 Female Tennis
11 Male Football
12 Female Tennis
13 Female Tennis
14 Male Football
15 Male Football
16 Male Basketball
17 Male Football
18 Female Basketball
19 Male Football
20 Male Football
21 Female Tennis
22 Male Football
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG US
10:20 30-12-2025

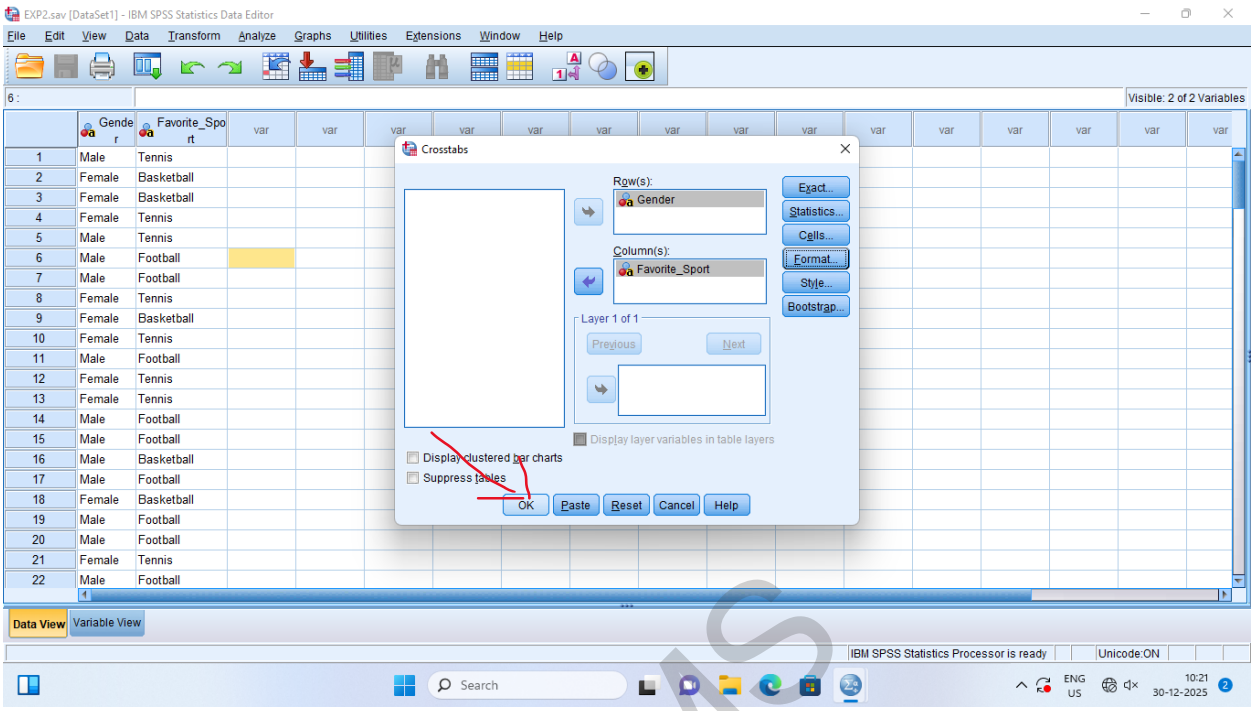

EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Reports
Descriptive Statistics
Bayesian Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Simulation...
Quality Control
ROC Curve...
Spatial and Temporal Modeling...
Direct Marketing
Frequencies...
Descriptives...
Explore...
Crosstabs...
TURF Analysis
Ratio...
P-P Plots...
Q-Q Plots...
Data View
Variable View
Crosstabs...
IBM SPSS Statistics Processor is ready
Unicode:ON


EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 2 of 2 Variables
Crosstabs
Row(s):
Gender
Column(s):
Favorite_Sport
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
OK Paste Reset Cancel Help
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON


EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1
Continue Cancel Help
OK Paste Reset Cancel Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON


EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 2 of 2 Variables
Gender
Favorite_Sport
Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than 5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:21
30-12-2025


EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 2 of 2 Variables
Gender
Favorite_Sport
Crosstabs
Row(s):
Gender
Column(s):
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
Crosstabs: Table Format
Row Order
Ascending
Descending
Continue
Cancel
Help
Display layer variables in table layers
Display clustered bar charts
Suppress tables
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:21
30-12-2025


EXP2.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help
Visible: 2 of 2 Variables
Crosstabs
Row(s):
Gender
Column(s):
Favorite_Sport
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON
Search
ENG
US
10:21
30-12-2025

Output:

Your temporary usage period for IBM SPSS Statistics will expire in 3653 days.

```
GET DATA
  /TYPE=XLSX
  /FILE='D:\palab\c_independence.xlsx'
  /SHEET=name 'c_independence'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.

DATASET NAME DataSet1 WINDOW=FRONT.
CROSSTABS
  /TABLES=Gender BY Favorite_Sport
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT EXPECTED ROW COLUMN TOTAL
  /COUNT ROUND CELL.
```


Crosstabs

[DataSet1]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Favorite_Sport	150	100.0%	0	0.0%	150	100.0%

Gender * Favorite_Sport Crosstabulation

			Favorite_Sport			
			Basketball	Football	Tennis	Total
Gender	Female	Count	15	5	50	70
		Expected Count	14.0	27.5	28.5	70.0
		% within Gender	21.4%	7.1%	71.4%	100.0%
		% within Favorite_Sport	50.0%	8.5%	82.0%	46.7%
		% of Total	10.0%	3.3%	33.3%	46.7%
	Male	Count	15	54	11	80
		Expected Count	16.0	31.5	32.5	80.0
		% within Gender	18.8%	67.5%	13.8%	100.0%
		% within Favorite_Sport	50.0%	91.5%	18.0%	53.3%
		% of Total	10.0%	36.0%	7.3%	53.3%
Total		Count	30	59	61	150
		Expected Count	30.0	59.0	61.0	150.0
		% within Gender	20.0%	39.3%	40.7%	100.0%
		% within Favorite_Sport	100.0%	100.0%	100.0%	100.0%
		% of Total	20.0%	39.3%	40.7%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	65.253[a]	2	.000
Likelihood Ratio	73.873	2	.000
N of Valid Cases	150		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.00.

Symmetric Measures

			Value	Approximate Significance
Nominal	by	Phi	.660	.000
Nominal		Cramer's V	.660	.000
N of Valid Cases			150	

Results

The two experiments were carried out successfully:

1. Chi-Square Goodness-of-Fit Test on the dice roll data.
2. Chi-Square Test of Independence on the Gender and Favorite_Sport data.

Exploratory Factor Analysis (EFA)

AIM:

To perform Exploratory Factor Analysis (EFA) using SPSS in order to identify the underlying factors among a set of observed variables and to reduce the data into a smaller number of meaningful latent constructs.

DESCRIPTION:

Exploratory Factor Analysis (EFA) is a statistical method used in multivariate data analysis to identify the underlying structure among a set of observed variables. It aims to reduce a large number of variables into a smaller set of latent (unobserved) factors that explain the patterns of correlations or covariances in the data. EFA is "exploratory" because it doesn't start with a predefined model; instead, it lets the data reveal the factor structure. It's commonly used in fields like psychology, social sciences, education, and marketing to develop or validate questionnaires, scales, or tests by grouping related items.

STEPS in EFA

In EFA, several key actions (or steps/options) are performed to extract, refine, and interpret factors.

1.Data Preparation:

Definition: Cleaning and selecting variables for analysis.

Explanation: Ensure no missing values (or handle them, e.g., via mean imputation or listwise deletion). Select only relevant variables (e.g., exclude IDs or demographics). Check assumptions like normality (via histograms or tests) and correlations (matrix should show patterns). In SPSS, this is done before running the analysis.

2.Suitability Testing:

Definition: Assessing if the data is appropriate for EFA.

Explanation: Use KMO (Kaiser-Meyer-Olkin) measure (>0.6 indicates adequate sampling) and Bartlett's Test of Sphericity ($p < 0.05$ means correlations are not zero). This prevents running EFA on unfactorable data.

3. **Factor Extraction**:

Definition: Methods to identify initial factors from the correlation matrix.

Explanation: Common methods include Principal Component Analysis (PCA, maximizes total variance; often used for data reduction) or Principal Axis Factoring (PAF, focuses on common variance; better for latent constructs). Extract factors based on criteria like eigenvalues >1 (Kaiser rule) or scree plot (visual "elbow" where variance drops).

4.**Factor Rotation**:

Definition: Adjusting factors to make the structure clearer and more interpretable.

Explanation: Unrotated factors can be hard to interpret due to mixed loadings. Orthogonal rotations (e.g., Varimax) assume independent factors and maximize high/low loadings. Oblique rotations (e.g., Promax) allow correlated factors, which is realistic in many cases (e.g., personality traits overlap). Rotation doesn't change the data but redistributes variance for simplicity.

5. **Factor Retention**:

Definition: Deciding how many factors to keep.

Explanation: Use multiple criteria: eigenvalues >1, scree plot, parallel analysis (compares to random data), or variance explained (>60% cumulative). Over-retaining leads to noise; under-retaining loses information.

6. **Item Refinement**:

Definition: Removing or adjusting problematic variables.

Explanation: Based on communalities (<0.3-0.5 means low explained variance), loadings (<0.4-0.5 or cross-loadings >0.3), or reliability (e.g., Cronbach's alpha per factor >0.7). Rerun EFA after removal for cleaner structure.

7.**Scoring and Further Analysis**:

Definition: Computing factor scores for use in other analyses.

Explanation: Save scores (e.g., regression method) for each case, then use in regression, clustering, or validation. This turns factors into usable variables.

Step-by-Step Process:

1. Open your dataset in **SPSS** using **File → Open → Data** and load the required data file.
2. Go to **Analyze → Dimension Reduction → Factor** to begin Exploratory Factor Analysis.
3. In the **Factor Analysis** dialog box, move the selected variables (x1 to x9) into the **Variables** box.
4. Click on **Descriptives**, select Univariate descriptive, coefficients, **KMO and Bartlett's Test of Sphericity** and **Anti-image correlations**, then click **Continue**.
5. Click on **Extraction**, choose **Principal Component** as the extraction method, select **Eigen values greater than 1**, and select correlation matrix, and check **Scree plot**, then click **Continue**.
6. Click on **Rotation**, select **Varimax rotation**, check **Rotated solution**, and click **Continue**.
7. Click on **Scores**, select **Save as variables** using the **Regression method**, and click **Continue**.
8. Click options, select exclude cases listwise and select "sorted by size",and click continue.
9. Click **OK** to run the Exploratory Factor Analysis.
10. Examine the output, including the **KMO and Bartlett's Test**, **Total Variance Explained**, **Scree Plot**, and **Rotated Factor Matrix** to determine the number of factors and item loadings.

```
GET DATA
  /TYPE=XLSX
  /FILE='C:\Users\sitams.LAB6-06\Downloads\bfi_efa.xlsx'
  /SHEET=name 'bfi'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet6 WINDOW=FRONT.
FACTOR
  /VARIABLES A1 A2 A3 A4 A5 C1 C2 C3 C4 C5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1 O2
 O3 O4 O5
  /MISSING LISTWISE
  /ANALYSIS A1 A2 A3 A4 A5 C1 C2 C3 C4 C5 E1 E2 E3 E4 E5 N1 N2 N3 N4 N5 O1 O2
O3 O4 O5
  /PRINT UNIVARIATE INITIAL CORRELATION KMO AIC EXTRACTION ROTATION
  /FORMAT SORT BLANK(.10)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /SAVE REG(ALL)
  /METHOD=CORRELATION.
```


Factor Analysis

```
[DataSet6]
```

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
A1	2.41	1.407	2436
A2	4.80	1.180	2436
A3	4.60	1.311	2436
A4	4.69	1.485	2436
A5	4.54	1.271	2436
C1	4.53	1.235	2436
C2	4.37	1.319	2436
C3	4.30	1.291	2436
C4	2.55	1.377	2436
C5	3.31	1.633	2436
E1	2.98	1.631	2436
E2	3.15	1.614	2436
E3	3.98	1.352	2436
E4	4.41	1.467	2436
E5	4.39	1.343	2436
N1	2.94	1.576	2436
N2	3.52	1.533	2436
N3	3.22	1.595	2436
N4	3.20	1.570	2436
N5	2.97	1.623	2436
O1	4.81	1.127	2436
O2	2.68	1.553	2436
O3	4.45	1.205	2436
O4	4.93	1.193	2436
O5	2.47	1.324	2436

Correlation Matrix

		A1	A2	A3	A4	A5	C1	C2
Correlation	A1	1.000	-.351	-.274	-.157	-.193	.015	.013
	A2	-.351	1.000	.503	.351	.397	.103	.130
	A3	-.274	.503	1.000	.385	.516	.114	.147
	A4	-.157	.351	.385	1.000	.326	.095	.226
	A5	-.193	.397	.516	.326	1.000	.135	.117
	C1	.015	.103	.114	.095	.135	1.000	.438
	C2	.013	.130	.147	.226	.117	.438	1.000
	C3	-.021	.189	.129	.133	.131	.317	.363
	C4	.115	-.146	-.121	-.172	-.126	-.362	-.396
	C5	.038	-.121	-.154	-.250	-.167	-.261	-.304
	E1	.107	-.222	-.210	-.134	-.252	-.031	.017
	E2	.088	-.243	-.292	-.210	-.338	-.107	-.075
	E3	-.049	.255	.383	.204	.412	.135	.153
	E4	-.070	.297	.388	.319	.483	.153	.123
	E5	-.020	.294	.253	.169	.269	.268	.258
	N1	.168	-.094	-.083	-.107	-.205	-.072	-.020
	N2	.140	-.050	-.092	-.155	-.199	-.038	-.006
	N3	.092	-.040	-.039	-.074	-.138	-.025	.003
	N4	.042	-.087	-.127	-.171	-.215	-.098	-.044
	N5	.015	.020	-.040	-.015	-.081	-.048	.051
	O1	.006	.126	.150	.057	.163	.179	.161
	O2	.066	.014	.007	.038	-.007	-.129	-.057
	O3	-.064	.165	.226	.071	.238	.197	.193
	O4	-.091	.083	.032	-.050	.011	.100	.047
	O5	.099	-.081	-.042	.027	-.050	-.130	-.067

Correlation Matrix

		C3	C4	C5	E1	E2	E3	E4
Correlation	A1	-.021	.115	.038	.107	.088	-.049	-.070
	A2	.189	-.146	-.121	-.222	-.243	.255	.297
	A3	.129	-.121	-.154	-.210	-.292	.383	.388
	A4	.133	-.172	-.250	-.134	-.210	.204	.319
	A5	.131	-.126	-.167	-.252	-.338	.412	.483
	C1	.317	-.362	-.261	-.031	-.107	.135	.153
	C2	.363	-.396	-.304	.017	-.075	.153	.123
	C3	1.000	-.358	-.351	-.009	-.089	.093	.100
	C4	-.358	1.000	.488	.098	.208	-.085	-.113
	C5	-.351	.488	1.000	.068	.266	-.163	-.207
	E1	-.009	.098	.068	1.000	.469	-.323	-.420
	E2	-.089	.208	.266	.469	1.000	-.386	-.527
	E3	.093	-.085	-.163	-.323	-.386	1.000	.417
	E4	.100	-.113	-.207	-.420	-.527	.417	1.000
	E5	.207	-.235	-.235	-.307	-.383	.391	.323
	N1	-.079	.216	.216	.015	.176	-.047	-.144
	N2	-.067	.158	.246	.020	.207	-.060	-.152
	N3	-.077	.202	.242	.055	.193	-.017	-.123
	N4	-.122	.271	.355	.235	.351	-.147	-.306
	N5	-.024	.197	.179	.054	.259	-.080	-.104
	O1	.091	-.094	-.086	-.105	-.158	.328	.138
	O2	-.030	.208	.123	.049	.079	-.074	.050
	O3	.059	-.083	-.074	-.217	-.238	.406	.215
	O4	.011	.053	.135	.094	.172	.053	-.099
	O5	-.003	.190	.056	.089	.081	-.123	.047

Correlation Matrix

		E5	N1	N2	N3	N4	N5	O1
Correlation	A1	-.020	.168	.140	.092	.042	.015	.006
	A2	.294	-.094	-.050	-.040	-.087	.020	.126
	A3	.253	-.083	-.092	-.039	-.127	-.040	.150
	A4	.169	-.107	-.155	-.074	-.171	-.015	.057
	A5	.269	-.205	-.199	-.138	-.215	-.081	.163
	C1	.268	-.072	-.038	-.025	-.098	-.048	.179
	C2	.258	-.020	-.006	.003	-.044	.051	.161
	C3	.207	-.079	-.067	-.077	-.122	-.024	.091
	C4	-.235	.216	.158	.202	.271	.197	-.094
	C5	-.235	.216	.246	.242	.355	.179	-.086
	E1	-.307	.015	.020	.055	.235	.054	-.105
	E2	-.383	.176	.207	.193	.351	.259	-.158
	E3	.391	-.047	-.060	-.017	-.147	-.080	.328
	E4	.323	-.144	-.152	-.123	-.306	-.104	.138
	E5	1.000	.034	.039	-.065	-.207	-.142	.291
	N1	.034	1.000	.718	.567	.410	.381	-.051
	N2	.039	.718	1.000	.550	.395	.351	-.043
	N3	-.065	.567	.550	1.000	.523	.431	-.040
	N4	-.207	.410	.395	.523	1.000	.403	-.049
	N5	-.142	.381	.351	.431	.403	1.000	-.129
	O1	.291	-.051	-.043	-.040	-.049	-.129	1.000
	O2	-.085	.135	.117	.104	.074	.187	-.230
	O3	.306	-.039	-.030	-.028	-.063	-.080	.393
	O4	-.001	.079	.131	.166	.220	.111	.177
	O5	-.113	.106	.024	.057	.036	.138	-.246

Correlation Matrix

		O2	O3	O4	O5
Correlation	A1	.066	-.064	-.091	.099
	A2	.014	.165	.083	-.081
	A3	.007	.226	.032	-.042
	A4	.038	.071	-.050	.027
	A5	-.007	.238	.011	-.050
	C1	-.129	.197	.100	-.130
	C2	-.057	.193	.047	-.067
	C3	-.030	.059	.011	-.003
	C4	.208	-.083	.053	.190
	C5	.123	-.074	.135	.056
	E1	.049	-.217	.094	.089
	E2	.079	-.238	.172	.081
	E3	-.074	.406	.053	-.123
	E4	.050	.215	-.099	.047
	E5	-.085	.306	-.001	-.113
	N1	.135	-.039	.079	.106
	N2	.117	-.030	.131	.024
	N3	.104	-.028	.166	.057
	N4	.074	-.063	.220	.036
	N5	.187	-.080	.111	.138
	O1	-.230	.393	.177	-.246
	O2	1.000	-.284	-.075	.328
	O3	-.284	1.000	.184	-.310
	O4	-.075	.184	1.000	-.182
	O5	.328	-.310	-.182	1.000

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.849
Bartlett's Test of Sphericity	Approx. Chi-Square	18146.066
	df	300
	Sig.	.000

Anti-image Matrices

		A1	A2	A3	A4	A5	C1
Anti-image Covariance	A1	.799	.168	.089	.019	.021	-.035
	A2	.168	.607	-.157	-.106	-.060	.015
	A3	.089	-.157	.567	-.102	-.148	-.003
	A4	.019	-.106	-.102	.730	-.044	.035
	A5	.021	-.060	-.148	-.044	.582	-.015
	C1	-.035	.015	-.003	.035	-.015	.706
	C2	-.034	.016	-.011	-.101	.018	-.173
	C3	-.028	-.078	-.001	.026	-.018	-.090
	C4	-.060	.010	-.002	.013	-.002	.098
	C5	.013	-.020	.005	.075	-.012	.022
	E1	-.045	.032	-.020	-.007	-.001	-.025
	E2	-.005	.017	.015	-.008	.005	-.017
	E3	-.036	.014	-.079	-.001	-.087	.022
	E4	-.036	-.013	-.040	-.068	-.121	-.051
	E5	-.034	-.095	.009	.010	-.015	-.070
	N1	-.037	.033	-.018	-.022	.034	.017
	N2	-.026	-.029	.002	.050	.023	.000
	N3	-.015	.006	-.015	-.020	.009	-.029
	N4	.026	-.021	-.004	.028	.010	.008
	N5	.026	-.046	.021	-.021	-.022	-.002
	O1	-.047	-.010	-.004	.004	-.002	-.023
	O2	-.036	-.031	-.015	-.019	-.015	.028
	O3	.006	.007	-.030	.029	-.024	-.033
	O4	.067	-.043	.003	.017	-.015	-.065
	O5	-.032	.028	-.014	-.020	.003	.026
Anti-image Correlation	A1	.754[a]	.241	.132	.024	.031	-.046
	A2	.241	.836[a]	-.267	-.159	-.101	.023
	A3	.132	-.267	.870[a]	-.159	-.258	-.004
	A4	.024	-.159	-.159	.878[a]	-.068	.048
	A5	.031	-.101	-.258	-.068	.904[a]	-.024
	C1	-.046	.023	-.004	.048	-.024	.843[a]
	C2	-.047	.025	-.018	-.147	.029	-.257
	C3	-.036	-.117	-.002	.036	-.027	-.124
	C4	-.088	.016	-.004	.019	-.004	.152
	C5	.019	-.033	.009	.112	-.019	.034

Anti-image Matrices

		C2	C3	C4	C5	E1	E2
Anti-image Covariance	A1	-.034	-.028	-.060	.013	-.045	-.005
	A2	.016	-.078	.010	-.020	.032	.017
	A3	-.011	-.001	-.002	.005	-.020	.015
	A4	-.101	.026	.013	.075	-.007	-.008
	A5	.018	-.018	-.002	-.012	-.001	.005
	C1	-.173	-.090	.098	.022	-.025	-.017
	C2	.642	-.111	.137	.042	-.061	-.020
	C3	-.111	.745	.092	.118	-.017	-.015
	C4	.137	.092	.585	-.174	-.038	-.027
	C5	.042	.118	-.174	.624	.058	-.048
	E1	-.061	-.017	-.038	.058	.668	-.151
	E2	-.020	-.015	-.027	-.048	-.151	.542
	E3	-.015	.005	-.031	.034	.055	.054
	E4	-.009	.015	-.030	.013	.108	.143
	E5	-.058	-.033	.023	.028	.064	.080
	N1	.003	.001	-.035	.016	.024	.008
	N2	-.006	.002	.029	-.044	.018	-.034
	N3	-.004	.008	-.010	-.013	.008	.011
	N4	-.029	.004	-.033	-.089	-.074	-.028
	N5	-.073	-.020	-.049	.014	.059	-.082
	O1	-.028	-.015	-.022	-.006	-.019	-.002
	O2	-.030	-.020	-.079	-.033	-.018	-.007
	O3	-.073	.018	-.051	-.017	.041	.026
	O4	.005	-.019	-.018	-.050	-.043	-.080
	O5	-.015	-.051	-.082	.006	-.040	-.017
Anti-image Correlation	A1	-.047	-.036	-.088	.019	-.062	-.007
	A2	.025	-.117	.016	-.033	.051	.029
	A3	-.018	-.002	-.004	.009	-.033	.027
	A4	-.147	.036	.019	.112	-.010	-.013
	A5	.029	-.027	-.004	-.019	-.002	.010
	C1	-.257	-.124	.152	.034	-.036	-.027
	C2	.796[a]	-.160	.223	.066	-.093	-.034
	C3	-.160	.852[a]	.139	.173	-.024	-.023
	C4	.223	.139	.827[a]	-.287	-.062	-.047
	C5	.066	.173	-.287	.864[a]	.091	-.082

Anti-image Matrices

		E3	E4	E5	N1	N2	N3
Anti-image Covariance	A1	-.036	-.036	-.034	-.037	-.026	-.015
	A2	.014	-.013	-.095	.033	-.029	.006
	A3	-.079	-.040	.009	-.018	.002	-.015
	A4	-.001	-.068	.010	-.022	.050	-.020
	A5	-.087	-.121	-.015	.034	.023	.009
	C1	.022	-.051	-.070	.017	.000	-.029
	C2	-.015	-.009	-.058	.003	-.006	-.004
	C3	.005	.015	-.033	.001	.002	.008
	C4	-.031	-.030	.023	-.035	.029	-.010
	C5	.034	.013	.028	.016	-.044	-.013
	E1	.055	.108	.064	.024	.018	.008
	E2	.054	.143	.080	.008	-.034	.011
	E3	.602	-.062	-.085	.000	.008	-.034
	E4	-.062	.528	-.011	.019	-.005	-.012
	E5	-.085	-.011	.633	-.054	-.048	.027
	N1	.000	.019	-.054	.410	-.230	-.099
	N2	.008	-.005	-.048	-.230	.428	-.092
	N3	-.034	-.012	.027	-.099	-.092	.524
	N4	-.003	.055	.039	-.039	-.013	-.157
	N5	.000	-.015	.062	-.054	-.024	-.099
	O1	-.099	-.003	-.085	-.001	.012	.005
	O2	-.015	-.051	-.008	-.009	-.024	.006
	O3	-.119	-.019	-.048	-.009	.004	.013
	O4	-.023	.013	.009	.020	-.023	-.036
	O5	.011	-.075	-.007	-.043	.039	-5.371E-5
Anti-image Correlation	A1	-.052	-.055	-.047	-.064	-.044	-.023
	A2	.023	-.023	-.154	.066	-.057	.010
	A3	-.136	-.072	.016	-.038	.004	-.028
	A4	-.002	-.109	.015	-.039	.090	-.033
	A5	-.146	-.218	-.025	.069	.045	.017
	C1	.034	-.084	-.105	.031	.001	-.048
	C2	-.025	-.016	-.091	.006	-.010	-.007
	C3	.008	.023	-.048	.002	.004	.013
	C4	-.052	-.055	.038	-.071	.059	-.017
	C5	.056	.022	.045	.031	-.085	-.023

Anti-image Matrices

		N4	N5	O1	O2	O3	O4
Anti-image Covariance	A1	.026	.026	-.047	-.036	.006	.067
	A2	-.021	-.046	-.010	-.031	.007	-.043
	A3	-.004	.021	-.004	-.015	-.030	.003
	A4	.028	-.021	.004	-.019	.029	.017
	A5	.010	-.022	-.002	-.015	-.024	-.015
	C1	.008	-.002	-.023	.028	-.033	-.065
	C2	-.029	-.073	-.028	-.030	-.073	.005
	C3	.004	-.020	-.015	-.020	.018	-.019
	C4	-.033	-.049	-.022	-.079	-.051	-.018
	C5	-.089	.014	-.006	-.033	-.017	-.050
	E1	-.074	.059	-.019	-.018	.041	-.043
	E2	-.028	-.082	-.002	-.007	.026	-.080
	E3	-.003	.000	-.099	-.015	-.119	-.023
	E4	.055	-.015	-.003	-.051	-.019	.013
	E5	.039	.062	-.085	-.008	-.048	.009
	N1	-.039	-.054	-.001	-.009	-.009	.020
	N2	-.013	-.024	.012	-.024	.004	-.023
	N3	-.157	-.099	.005	.006	.013	-.036
	N4	.561	-.101	-.021	.014	-.008	-.059
	N5	-.101	.682	.050	-.059	-.006	-.016
	O1	-.021	.050	.744	.083	-.125	-.086
	O2	.014	-.059	.083	.786	.133	.003
	O3	-.008	-.006	-.125	.133	.647	-.080
	O4	-.059	-.016	-.086	.003	-.080	.822
	O5	.007	-.043	.073	-.145	.120	.090
Anti-image Correlation	A1	.038	.036	-.060	-.045	.008	.083
	A2	-.035	-.071	-.015	-.045	.011	-.061
	A3	-.006	.033	-.006	-.023	-.049	.004
	A4	.044	-.030	.006	-.025	.043	.022
	A5	.018	-.034	-.003	-.022	-.039	-.022
	C1	.012	-.002	-.032	.037	-.050	-.085
	C2	-.049	-.110	-.041	-.042	-.113	.007
	C3	.006	-.028	-.020	-.027	.027	-.024
	C4	-.058	-.077	-.034	-.116	-.083	-.027
	C5	-.150	.021	-.008	-.048	-.027	-.069

Anti-image Matrices

		O5
Anti-image Covariance	A1	-.032
	A2	.028
	A3	-.014
	A4	-.020
	A5	.003
	C1	.026
	C2	-.015
	C3	-.051
	C4	-.082
	C5	.006
	E1	-.040
	E2	-.017
	E3	.011
	E4	-.075
	E5	-.007
	N1	-.043
	N2	.039
	N3	-5.371E-5
	N4	.007
	N5	-.043
	O1	.073
	O2	-.145
	O3	.120
	O4	.090
	O5	.768
Anti-image Correlation	A1	-.041
	A2	.041
	A3	-.020
	A4	-.026
	A5	.005
	C1	.035
	C2	-.021
	C3	-.067
	C4	-.122
	C5	.008

Anti-image Matrices

		A1	A2	A3	A4	A5	C1
	E1	-.062	.051	-.033	-.010	-.002	-.036
	E2	-.007	.029	.027	-.013	.010	-.027
	E3	-.052	.023	-.136	-.002	-.146	.034
	E4	-.055	-.023	-.072	-.109	-.218	-.084
	E5	-.047	-.154	.016	.015	-.025	-.105
	N1	-.064	.066	-.038	-.039	.069	.031
	N2	-.044	-.057	.004	.090	.045	.001
	N3	-.023	.010	-.028	-.033	.017	-.048
	N4	.038	-.035	-.006	.044	.018	.012
	N5	.036	-.071	.033	-.030	-.034	-.002
	O1	-.060	-.015	-.006	.006	-.003	-.032
	O2	-.045	-.045	-.023	-.025	-.022	.037
	O3	.008	.011	-.049	.043	-.039	-.050
	O4	.083	-.061	.004	.022	-.022	-.085
	O5	-.041	.041	-.020	-.026	.005	.035

Anti-image Matrices

		C2	C3	C4	C5	E1	E2
	E1	-.093	-.024	-.062	.091	.838[a]	-.251
	E2	-.034	-.023	-.047	-.082	-.251	.884[a]
	E3	-.025	.008	-.052	.056	.087	.095
	E4	-.016	.023	-.055	.022	.182	.268
	E5	-.091	-.048	.038	.045	.098	.137
	N1	.006	.002	-.071	.031	.045	.017
	N2	-.010	.004	.059	-.085	.034	-.071
	N3	-.007	.013	-.017	-.023	.014	.020
	N4	-.049	.006	-.058	-.150	-.120	-.051
	N5	-.110	-.028	-.077	.021	.087	-.134
	O1	-.041	-.020	-.034	-.008	-.027	-.003
	O2	-.042	-.027	-.116	-.048	-.024	-.011
	O3	-.113	.027	-.083	-.027	.062	.045
	O4	.007	-.024	-.027	-.069	-.058	-.120
	O5	-.021	-.067	-.122	.008	-.056	-.026

Anti-image Matrices

	E3	E4	E5	N1	N2	N3
E1	.087	.182	.098	.045	.034	.014
E2	.095	.268	.137	.017	-.071	.020
E3	.897[a]	-.110	-.138	.000	.015	-.060
E4	-.110	.877[a]	-.019	.041	-.010	-.023
E5	-.138	-.019	.893[a]	-.105	-.092	.046
N1	.000	.041	-.105	.779[a]	-.548	-.214
N2	.015	-.010	-.092	-.548	.780[a]	-.194
N3	-.060	-.023	.046	-.214	-.194	.862[a]
N4	-.005	.101	.066	-.082	-.026	-.290
N5	.000	-.026	.094	-.103	-.044	-.166
O1	-.148	-.006	-.125	-.001	.022	.008
O2	-.022	-.080	-.012	-.016	-.042	.009
O3	-.191	-.032	-.076	-.017	.007	.022
O4	-.033	.019	.012	.034	-.039	-.055
O5	.017	-.117	-.011	-.077	.069	-8.467E-5

Anti-image Matrices

	N4	N5	O1	O2	O3	O4
E1	-.120	.087	-.027	-.024	.062	-.058
E2	-.051	-.134	-.003	-.011	.045	-.120
E3	-.005	.000	-.148	-.022	-.191	-.033
E4	.101	-.026	-.006	-.080	-.032	.019
E5	.066	.094	-.125	-.012	-.076	.012
N1	-.082	-.103	-.001	-.016	-.017	.034
N2	-.026	-.044	.022	-.042	.007	-.039
N3	-.290	-.166	.008	.009	.022	-.055
N4	.885[a]	-.163	-.032	.021	-.014	-.086
N5	-.163	.860[a]	.070	-.081	-.009	-.022
O1	-.032	.070	.859[a]	.108	-.181	-.110
O2	.021	-.081	.108	.780[a]	.187	.004
O3	-.014	-.009	-.181	.187	.844[a]	-.110
O4	-.086	-.022	-.110	.004	-.110	.770[a]
O5	.011	-.060	.096	-.187	.170	.114

Anti-image Matrices

	O5
E1	-.056
E2	-.026
E3	.017
E4	-.117
E5	-.011
N1	-.077
N2	.069
N3	-8.467E-5
N4	.011
N5	-.060
O1	.096
O2	-.187
O3	.170
O4	.114
O5	.762[a]

a. Measures of Sampling Adequacy(MSA)

Communalities

	Initial	Extraction
A1	1.000	.658
A2	1.000	.610
A3	1.000	.608
A4	1.000	.426
A5	1.000	.579
C1	1.000	.500
C2	1.000	.603
C3	1.000	.479
C4	1.000	.672
C5	1.000	.545
E1	1.000	.604
E2	1.000	.637
E3	1.000	.584
E4	1.000	.624
E5	1.000	.509
N1	1.000	.745
N2	1.000	.727
N3	1.000	.645
N4	1.000	.588
N5	1.000	.482
O1	1.000	.533
O2	1.000	.515
O3	1.000	.580
O4	1.000	.486
O5	1.000	.563

Extraction Method: Principal Component Analysis.

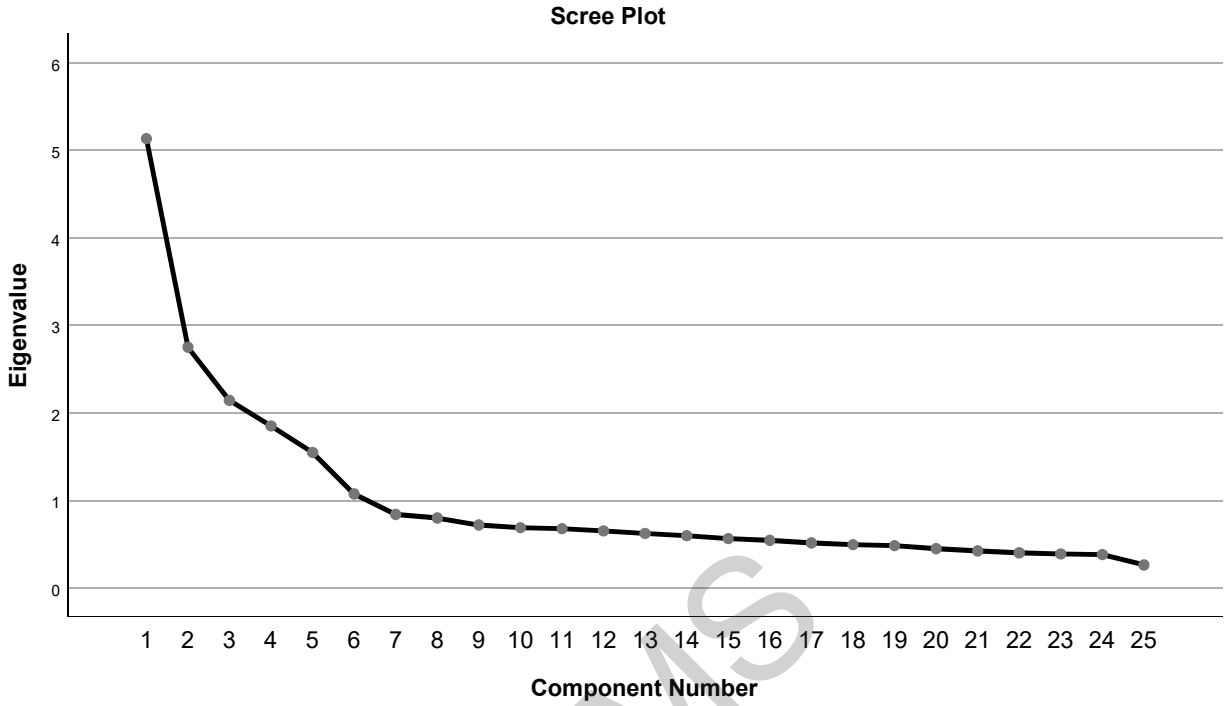
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.134	20.537	20.537	5.134	20.537	20.537
2	2.752	11.008	31.545	2.752	11.008	31.545
3	2.143	8.571	40.116	2.143	8.571	40.116
4	1.852	7.409	47.525	1.852	7.409	47.525
5	1.548	6.193	53.718	1.548	6.193	53.718
6	1.074	4.294	58.012	1.074	4.294	58.012
7	.840	3.358	61.370			
8	.799	3.197	64.567			
9	.719	2.876	67.443			
10	.688	2.752	70.195			
11	.676	2.705	72.901			
12	.652	2.607	75.508			
13	.623	2.493	78.001			
14	.597	2.386	80.387			
15	.563	2.252	82.640			
16	.543	2.173	84.813			
17	.515	2.058	86.871			
18	.495	1.978	88.849			
19	.483	1.931	90.779			
20	.449	1.796	92.575			
21	.423	1.693	94.269			
22	.401	1.603	95.871			
23	.388	1.551	97.422			
24	.382	1.527	98.950			
25	.263	1.050	100.000			

Total Variance Explained

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.093	12.371	12.371
2	2.593	10.373	22.744
3	2.577	10.306	33.051
4	2.533	10.131	43.182
5	2.095	8.382	51.563
6	1.612	6.449	58.012
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			

Extraction Method: Principal Component Analysis.



Component Matrix[a]

	Component					
	1	2	3	4	5	6
E2	-.642		.259		.348	.172
E4	.622	.166	-.377		-.217	.116
A5	.613	.189	-.308		.188	.194
E3	.563	.360	-.136	-.207	-.155	.229
A3	.557	.316	-.295		.318	
E5	.553	.312			-.307	
N4	-.547	.481			.214	
C5	-.524	.190	-.294	-.351	.108	.117
A2	.496	.313	-.225	.123	.415	-.169
C4	-.493	.132	-.477	-.279		.326
A4	.457	.125	-.190	.328	.235	
E1	-.443	-.193	.327	.157	.336	.355
N3	-.415	.672				
N2	-.425	.653			-.229	-.237
N1	-.437	.652		.118	-.282	-.186
N5	-.369	.500		.271	.137	

Component Matrix[a]

	Component					
	1	2	3	4	5	6
C1	.376	.154	.523	.209		.130
C2	.361	.216	.506	.382		.154
O5	-.225		-.369	.507	-.150	.300
O3	.434	.331	.195	-.473		.140
O2	-.219		-.412	.459		.281
C3	.350		.394	.441		
O1	.361	.247	.271	-.420		.300
A1	-.250		.173		-.610	.437
O4		.340	.258	-.286	.414	.214

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

Rotated Component Matrix[a]

	Component					
	1	2	3	4	5	6
N1	.837		-.166			
N2	.835		-.151			
N3	.795					
N4	.617	-.171		.418		
N5	.608		.140	.216		.199
C2		.738			.201	
C4	.186	-.689		.162	.150	.325
C3		.678	.123			
C1		.653			.263	
C5	.284	-.625		.236		
A2		.120	.749	-.145		
A3			.710	-.195	.228	
A1			-.661		.224	.393
A5	-.190		.584	-.254	.328	.163
A4		.249	.548	-.128		.196
E1			-.189	.729		.154
E2	.214		-.187	.728	-.134	
E4	-.146		.352	-.582	.253	.261

Rotated Component Matrix[a]

	Component					
	1	2	3	4	5	6
E5		.333	.113	-.512	.335	
O4	.193		.170	.434	.430	-.215
O1		.116			.689	-.194
O3			.105	-.181	.663	-.305
E3			.261	-.416	.579	
O5					-.249	.704
O2	.149				-.184	.662

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

a. Rotation converged in 7 iterations.

Component Transformation Matrix

Component	1	2	3	4	5	6
1	-.442	.429	.477	-.513	.339	-.120
2	.845	.067	.301	-.117	.421	.011
3	.055	.682	-.408	.369	.187	-.441
4	.152	.584	.181	.043	-.530	.567
5	-.112	-.076	.683	.681	-.058	-.217
6	-.228	.006	-.123	.347	.624	.650

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

RESULT:

The dataset was successfully tested and found suitable for Exploratory Factor Analysis.Meaningful factors were successfully extracted and rotated for clear interpretation. The extracted factors were validated and confirmed to be reliable.

Cluster Analysis

Aim:

To apply cluster analysis on the Iris dataset to group the flower observations into 3 clusters based on their measurements (sepal length, sepal width, petal length, petal width), using unsupervised methods.

Description:

Definition of Cluster Analysis:

Cluster analysis is an unsupervised machine learning technique that partitions a dataset into groups (clusters) based on similarity in features, without relying on predefined labels. Similarities are typically measured using distance metrics like Euclidean distance.

Types of Cluster Analysis:

1. **Hierarchical Clustering**: An agglomerative (bottom-up) approach that starts with each data point as a single cluster and merges them iteratively based on linkage criteria (e.g., Ward's method minimizes variance). Produces a dendrogram for visualizing cluster hierarchy.
2. **K-Means Clustering**: A partitioning algorithm that assigns data points to a fixed number of clusters (k=3 here) by iteratively updating cluster centroids to minimize the sum of squared distances within clusters.
3. **Two-Step Clustering**: A hybrid scalable method that first pre-clusters data into many small sub-clusters (using log-likelihood or Euclidean distance), then applies hierarchical clustering to merge them. It can handle large datasets and mixed variable types.

Formula for Z-Score Standardization: To prevent variables with larger scales from dominating distance calculations, standardize each feature using Z-scores:

$$Z = (x - \mu) / \sigma$$

- **x**: Original value
- **μ**: Mean of the variable across all observations
- **σ**: Standard deviation of the variable This scales all variables to mean=0 and SD=1.

Data Preprocessing: Import and Standardize the Iris Dataset in SPSS

1. Open SPSS.
2. **Import Excel File**:
    - **File → Import Data → Excel**
    - Select "irisdataexcel.xlsx".
    - Choose Sheet1 (it contains the data).
    - Check **Read variable names from the first row**.
    - OK .
- **Define Variables**:Variable View → Set numeric variables to Scale; target to Nominal.
- **Standardize Variables**:
    - **Analyze → Descriptive Statistics → Descriptives**.
    - Move all 4 variables to Variable(s).
    - Check **Save standardized values as variables**.
    - OK → Creates Zpetal length (cm), Zpetal width (cm), etc. Use these for clustering.

Method 1: Hierarchical Cluster Analysis (For 3 Clusters)

Uses Ward's method for balanced clusters; dendrogram helps confirm 3 is suitable.

1. **Analyze → Classify → Hierarchical Cluster…**
2. **Variables**: Move Zpetal length (cm), Zpetal width (cm) to Variable(s). (Ignore target.)
3. **Cluster**: Cases (default).
4. Click **Method**: Cluster Method → Ward's method; Measure → Squared Euclidean distance. OK.
5. Click **Statistics**: Agglomeration schedule; Cluster Membership → Single solution → 3. OK.
6. Click **Plots**: Dendrogram. OK.
7. Click **Save**: Cluster membership → Single solution → 3 (creates new variable, e.g., CLU3_1). OK.
8. OK to run.
9. Interpret: Dendrogram shows clear 3-cluster cut; crosstab CLU3_1 vs. target for match (~83% with all features, higher with petals).

Method 2: K-Means Cluster Analysis (Force 3 Clusters)

Iterative partitioning; random initialization may vary results slightly—rerun if needed for best match.

1. **Analyze → Classify → K-Means Cluster…**
2. **Variables**: Move Zpetal length (cm), Zpetal width (cm) to Variable(s). (Ignore target.)
3. **Number of Clusters**: 3.
4. Click **Iterate**: Max iterations 10, Convergence 0 (default).
5. Click **Save**: Cluster membership (creates QCL_1).
6. Click **Options**: Initial cluster centers, ANOVA table. OK.
7. OK to run.
8. Interpret: Cluster Centers table shows group means (e.g., small petals = setosa); ANOVA identifies key variables. Crosstab QCL_1 vs. target (~96% match with petals).

```
Your temporary usage period for IBM SPSS Statistics will expire in 3625 days.

GET DATA
  /TYPE=XLSX
  /FILE='D:\palab\irisdataexcel.xlsx'
  /SHEET=name 'Sheet1'
  /CELLRANGE=FULL
  /READNAMES=ON
  /DATATYPEMIN PERCENTAGE=95.0
  /HIDDEN IGNORE=YES.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
DESCRIPTIVES VARIABLES=sepallengthcm sepalwidthcm petallengthcm petalwidthcm
  /SAVE
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

```
[DataSet1]
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
sepal length (cm)	150	4.3	7.9	5.843	.8281
sepal width (cm)	150	2.0	4.4	3.054	.4336
petal length (cm)	150	1.0	6.9	3.759	1.7644
petal width (cm)	150	.1	2.5	1.199	.7632
Valid N (listwise)	150				

```
CLUSTER   Zsepallengthcm Zsepalwidthcm Zpetallengthcm Zpetalwidthcm
  /METHOD WARD
  /MEASURE=SEUCLID
  /PRINT SCHEDULE CLUSTER(3)
  /PLOT DENDROGRAM VICICLE
  /SAVE CLUSTER(3).
```

Cluster

Case Processing Summary[a,b]

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
150	100.0	0	.0	150	100.0

a. Squared Euclidean Distance used

b. Ward Linkage

Ward Linkage

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	102	143	.000	0	0	68
2	35	38	.000	0	0	3
3	10	35	.000	0	2	25
4	8	40	.007	0	0	22
5	11	49	.015	0	0	97
6	129	133	.023	0	0	92
7	1	18	.032	0	0	15
8	128	139	.041	0	0	34
9	3	48	.050	0	0	16
10	81	82	.060	0	0	72
11	20	47	.070	0	0	50
12	2	26	.084	0	0	37
13	121	144	.097	0	0	70
14	12	25	.112	0	0	60
15	1	41	.127	7	0	29
16	3	30	.142	9	0	44
17	89	96	.160	0	0	71
18	137	149	.182	0	0	79
19	64	79	.204	0	0	39
20	14	39	.227	0	0	49
21	66	87	.250	0	0	81
22	8	29	.274	4	0	90
23	56	100	.299	0	0	38

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
24	6	17	.325	0	0	86
25	10	31	.351	3	0	82
26	83	93	.380	0	0	54
27	67	85	.409	0	0	106
28	75	98	.438	0	0	77
29	1	28	.467	15	0	52
30	36	50	.500	0	0	109
31	117	138	.534	0	0	59
32	58	94	.568	0	0	83
33	13	46	.603	0	0	37
34	128	150	.638	8	0	126
35	124	127	.673	0	0	64
36	113	140	.709	0	0	76
37	2	13	.745	12	33	82
38	56	97	.782	23	0	96
39	64	92	.823	19	0	67
40	21	32	.864	0	0	58
41	59	76	.906	0	0	85
42	70	90	.949	0	0	72
43	24	27	.993	0	0	74
44	3	4	1.039	16	0	62
45	108	131	1.088	0	0	114
46	91	95	1.136	0	0	94
47	52	57	1.185	0	0	103
48	51	53	1.234	0	0	81
49	9	14	1.284	0	20	119
50	20	45	1.336	11	0	53
51	142	146	1.393	0	0	57
52	1	5	1.451	29	0	90
53	20	22	1.510	50	0	97
54	68	83	1.570	0	26	75
55	55	134	1.639	0	0	100
56	69	120	1.708	0	0	88
57	141	142	1.782	0	51	101

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
58	21	37	1.857	40	0	111
59	117	148	1.934	31	0	78
60	7	12	2.010	0	14	102
61	118	132	2.088	0	0	123
62	3	43	2.170	44	0	119
63	106	136	2.251	0	0	117
64	112	124	2.337	0	35	113
65	72	74	2.424	0	0	77
66	111	116	2.515	0	0	110
67	62	64	2.608	0	39	104
68	102	122	2.702	1	0	98
69	33	34	2.804	0	0	112
70	121	145	2.910	13	0	80
71	65	89	3.015	0	17	96
72	70	81	3.124	42	10	87
73	84	135	3.232	0	0	113
74	24	44	3.340	43	0	109
75	68	80	3.452	54	0	115
76	103	113	3.565	0	36	101
77	72	75	3.682	65	28	104
78	104	117	3.799	0	59	93
79	101	137	3.916	0	18	132
80	121	125	4.037	70	0	110
81	51	66	4.162	48	21	136
82	2	10	4.291	37	25	125
83	58	99	4.423	32	0	121
84	73	147	4.562	0	0	105
85	59	77	4.703	41	0	100
86	6	19	4.846	24	0	99
87	54	70	4.990	0	72	128
88	69	88	5.137	56	0	108
89	126	130	5.284	0	0	114
90	1	8	5.436	52	22	111
91	71	86	5.599	0	0	103

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
92	105	129	5.762	0	6	120
93	78	104	5.933	0	78	120
94	60	91	6.121	0	46	115
95	119	123	6.311	0	0	117
96	56	65	6.509	38	71	106
97	11	20	6.713	5	53	124
98	102	114	6.937	68	0	107
99	6	15	7.175	86	0	124
100	55	59	7.424	55	85	129
101	103	141	7.677	76	57	122
102	7	23	7.954	60	0	127
103	52	71	8.240	47	91	126
104	62	72	8.561	67	77	129
105	73	109	8.882	84	0	118
106	56	67	9.215	96	27	140
107	102	115	9.567	98	0	134
108	63	69	9.940	0	88	138
109	24	36	10.314	74	30	116
110	111	121	10.718	66	80	122
111	1	21	11.130	90	58	116
112	16	33	11.542	0	69	135
113	84	112	11.972	73	64	118
114	108	126	12.405	45	89	131
115	60	68	12.864	94	75	128
116	1	24	13.400	111	109	127
117	106	119	13.952	63	95	131
118	73	84	14.508	105	113	134
119	3	9	15.129	62	49	125
120	78	105	15.755	93	92	137
121	58	61	16.417	83	0	133
122	103	111	17.119	101	110	132
123	110	118	17.837	0	61	143
124	6	11	18.678	99	97	135
125	2	3	19.522	82	119	141

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
126	52	128	20.410	103	34	137
127	1	7	21.306	116	102	141
128	54	60	22.246	87	115	130
129	55	62	23.270	100	104	136
130	54	107	24.442	128	0	138
131	106	108	25.801	117	114	144
132	101	103	27.427	79	122	143
133	42	58	29.443	0	121	145
134	73	102	31.465	118	107	142
135	6	16	33.740	124	112	146
136	51	55	36.028	81	129	139
137	52	78	38.672	126	120	139
138	54	63	42.299	130	108	140
139	51	52	46.386	136	137	142
140	54	56	52.330	138	106	145
141	1	2	60.030	127	125	146
142	51	73	67.768	139	134	147
143	101	110	75.562	132	123	144
144	101	106	84.476	143	131	147
145	42	54	93.443	133	140	148
146	1	6	117.242	141	135	149
147	51	101	149.122	142	144	148
148	42	51	228.621	145	147	149
149	1	42	596.000	146	148	0

Cluster Membership

Case	3 Clusters
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1

Cluster Membership

Case	3 Clusters
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	2
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	3
52	3
53	3
54	2
55	3
56	2
57	3
58	2
59	3
60	2
61	2
62	3
63	2
64	3
65	2
66	3
67	2
68	2

Cluster Membership

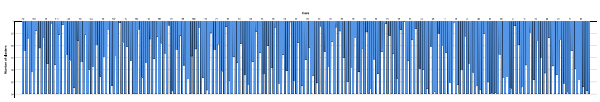
Case	3 Clusters
69	2
70	2
71	3
72	3
73	3
74	3
75	3
76	3
77	3
78	3
79	3
80	2
81	2
82	2
83	2
84	3
85	2
86	3
87	3
88	2
89	2
90	2
91	2
92	3
93	2
94	2
95	2
96	2
97	2
98	3
99	2
100	2
101	3
102	3

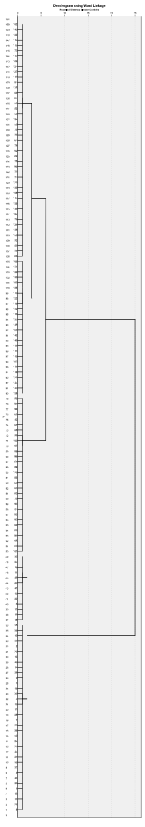
Cluster Membership

Case	3 Clusters
103	3
104	3
105	3
106	3
107	2
108	3
109	3
110	3
111	3
112	3
113	3
114	3
115	3
116	3
117	3
118	3
119	3
120	2
121	3
122	3
123	3
124	3
125	3
126	3
127	3
128	3
129	3
130	3
131	3
132	3
133	3
134	3
135	3
136	3

Cluster Membership

Case	3 Clusters
137	3
138	3
139	3
140	3
141	3
142	3
143	3
144	3
145	3
146	3
147	3
148	3
149	3
150	3





```
QUICK CLUSTER Zsepallengthcm Zsepalwidthcm Zpetallengthcm Zpetalwidthcm
  /MISSING=LISTWISE
  /CRITERIA=CLUSTER(3) MXITER(10) CONVERGE(0)
  /METHOD=KMEANS(NOUPDATE)
  /SAVE CLUSTER
  /PRINT INITIAL ANOVA.
```

Quick Cluster

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore: sepal length (cm)	2.24217	-1.62225	-.17309
Zscore: sepal width (cm)	-1.04706	-1.73895	3.10428
Zscore: petal length (cm)	1.78038	-1.39347	-1.28012
Zscore: petal width (cm)	1.44312	-1.17756	-1.04652

Iteration History[a]

	Change in Cluster Centers		
Iteration	1	2	3
1	1.985	1.786	1.811
2	.034	.281	.236
3	.018	.673	.297
4	.119	.258	.030
5	.122	.177	.000
6	.090	.100	.000
7	.039	.045	.000
8	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 8. The minimum distance between initial centers is 5.058.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore: sepal length (cm)	1.03015	-.16784	-.99872
Zscore: sepal width (cm)	.01384	-.96684	.89212
Zscore: petal length (cm)	.94054	.25875	-1.29862
Zscore: petal width (cm)	.96902	.17551	-1.25244

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore: sepal length (cm)	54.268	2	.275	147	197.153	.000
Zscore: sepal width (cm)	41.004	2	.456	147	89.975	.000
Zscore: petal length (cm)	67.185	2	.100	147	675.014	.000
Zscore: petal width (cm)	64.961	2	.130	147	500.559	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	55.000
	2	46.000
	3	49.000
Valid		150.000
Missing		.000

Exp 05

5. Logistic Regression

Aim:- To build and train a logistic regression model using spss 25.

Description:- Logistic regression is a supervised classification technique used to predict the probability of occurances of a categorical dependent variable. Typically binary in nature (such as Yes/No, 1/0).

→ It uses a sigmoid function to transform a linear combination of independent variable into a probability value blw 0 and 1.

Mathematically, logistic regression is expressed as:

$$z = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n.$$

$$P = \frac{1}{1+e^{-z}}$$

Where;

P = probability of occurence of the event,

b_0 = intercept

$b_1 b_2 \ldots b_n$ = Regression coefficients.

$x_1, x_2 \ldots x_n$ = independent variables.

If the predicted probability is greater than 0 or equal to 0.5, the outcome is classified as positive[1] otherwise, it is classified as negative[0]

Dataset Descriptions:-

Dependent variable:-

⇒ Disease. (0 = No Disease, 1 = Disease Present)

Independent variable:-

⇒ Age (in years)

⇒ BP - Blood pressure

⇒ sugar - Blood sugar level.

Sample future prediction (manual explanation):-

Assume spss estimates the following eqn

$z = -12 + 0.08(Age) + 0.04(BP) + 0.03(Sugar)$

for a new patient:

⇒ Age = 42

⇒ BP = 155

⇒ sugar = 140

$z = -12 + (0.08 \times 42) + (0.04 \times 155) + (0.03 \times 140)$

$z = 1.76$

$P = \frac{1}{1+e^{-1.76}} \approx 0.85$

Since, $0.85 \geq 0.5$, the patient is predicted to have disease (1).

Step-by-step process:-

1. Open spss 25. click file, select open, then data, choose excel file type, select the dataset, tick the option Read variable names from first raw, & click ok.
2. Go to Data view & check that all training data is entered correctly & that there are no missing values in the dependents variables of the training records.
3. click Analyze, select Regression, and then choose binary logistic to start model building.
4. Move disease into the dependent box and move Age, blood pressure, & sugar into the covariables box.
5. Click save and select probabilities and group membership, then click continue.
6. Click options, select Hosmer-Lemeshow goodness of

fit and classification plots and
continue,

* sig > 0.05 → Good model fit ☑
* sig < 0.05 → poor model fit ☒

7. click ok to run the logistic regression model. SPSS builds & trains the model using the existing dataset.

8. Analyze the o/p by checking the omni by test to confirm model significance, the Hosmer-Lemeshow test for model fit, the classification table for prediction accuracy and the variables in the Equation table for significant predictors.

9. To predict future outcomes go to data view and scroll to the bottom of the dataset → Add new rows and enter values only for the predictor variables, leaving the dependent variable disease blank.

10. Run the logistic regression again using the same settings SPSS applies the trained model to the new data.

11. Observe the newly created predicted probability and predicted group columns in data view to identify future predictions.

12. Interpret the results by classifying outcomes based on prediction probability. where values greater than or equal to 0.5 indicate presence of disease and values below 0.5 indicate absence of disease.

13. Conclude that logistic regression was successfully used to train a model and predict future disease outcomes using SPSS.

output:-

Logistic regression:-

Hosmer & Lemenshow Test

step	chi-square	df	sig
1	6.496	8	.592

Variables in the Eqn:-

	B	S.E	wald	df	Sig	Exp(B)
Age	.515	.269	3.657	1	.056	1.674
BP	-.038	.192	.038	1	.845	.963
Sugar	-.133	.127	1.091	1	.296	.876
Constant	.692	19.581	.001	1	.972	.1.997

a. variable(s) entered on step1: Age, BP, sugar

Result:-

Thus, the dataset building & travelling or training a logistic Regression model using SPSS 25 is successfully

Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	70	97.2
	Missing Cases	2	2.8
	Total	72	100.0
Unselected Cases		0	.0
Total		72	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			Disease		Percentage Correct
	Observed		0	1	
Step 0	Disease	0	37	0	100.0
		1	33	0	.0
	Overall Percentage				52.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.114	.239	.228	1	.633	.892

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Age	37.745	1	.000
		BP	33.096	1	.000
		Sugar	33.052	1	.000
	Overall Statistics		39.219	3	.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	52.176	3	.000
	Block	52.176	3	.000
	Model	52.176	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	44.636[a]	.525	.701

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.496	8	.592

Contingency Table for Hosmer and Lemeshow Test

		Disease = 0		Disease = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	7	6.887	0	.113	7
	2	7	6.730	0	.270	7
	3	7	6.433	0	.567	7
	4	6	5.952	1	1.048	7
	5	4	5.111	3	1.889	7
	6	2	3.380	5	3.620	7
	7	2	1.638	5	5.362	7
	8	2	.642	5	6.358	7
	9	0	.198	7	6.802	7
	10	0	.028	7	6.972	7

Classification Table[a]

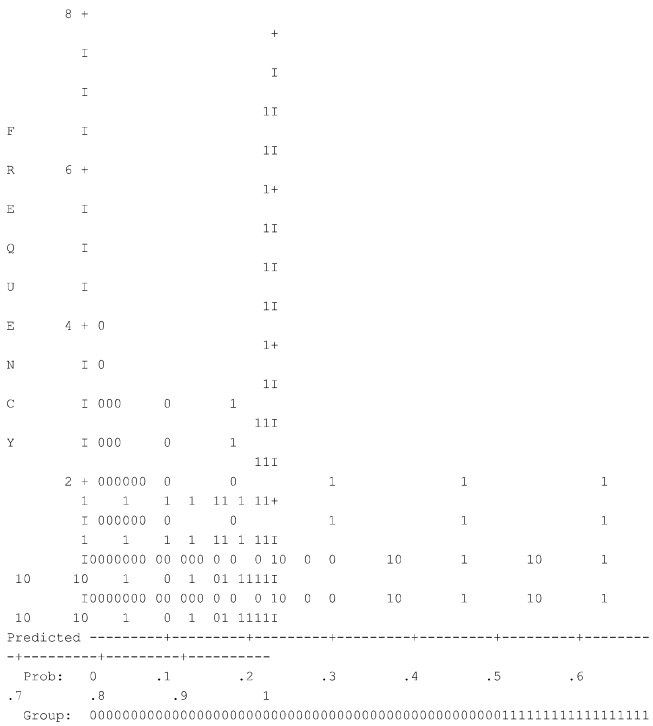
			Predicted		
			Disease		Percentage Correct
	Observed		0	1	
Step 1	Disease	0	32	5	86.5
		1	6	27	81.8
	Overall Percentage				84.3

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Age	.515	.269	3.657	1	.056	1.674
	BP	-.038	.192	.038	1	.845	.963
	Sugar	-.133	.127	1.091	1	.296	.876
	Constant	.692	19.581	.001	1	.972	1.997

a. Variable(s) entered on step 1: Age, BP, Sugar.



Step number: 1

Observed Groups and Predicted Probabilities

Predicted Probability is of Membership for 1

The Cut Value is .50

Symbols: 0 - 0
         1 - 1
Each Symbol Represents .5 Cases.