1 Data Analysis Software Basics

(PSY206) Data Management and Analysis

Md Rasel Biswas

Subsection 1

1.1 Word and Spreadsheet

Overview

- In data analysis, we often deal with **large amounts of text, numbers, and tables**.
- Two essential tools to manage these are the **word processor** and the **spreadsheet**.
- These software packages are foundational:
  - Word processors help us **create, format, and edit documents**.
  - Spreadsheets help us **organize, calculate, and analyze numerical data**.
- Before moving on to statistical software (like `SPSS`, `Nvivo`, or `MAXQDA`), students must have a clear understanding of these fundamental tools.

Word Processors

- A **word processor** is software used for creating, editing, formatting, and printing text-based documents.
- They replaced traditional typewriters by allowing:
  - Easy editing and revising of text.
  - Rich formatting options (fonts, margins, headings, alignment).
  - Insertion of **tables, figures, footnotes, references, hyperlinks**.
  - Spell-checking and grammar tools.
- Examples: **Microsoft Word, Google Docs, LibreOffice Writer, Apple Pages, WPS Writer, Overleaf (LaTeX editor)**.

Microsoft Word

- Part of the **Microsoft Office Suite**.
- **Features:**
  - Templates for reports, resumes, academic theses.
  - Advanced referencing tools (citations, bibliographies).
  - Track changes and comments for collaboration.
  - Mail merge for generating personalized letters.
- **Strengths:**
  - Professional, flexible, widely supported across industries.
- **Limitations:**
  - Paid software requiring license/subscription.

Microsoft Word

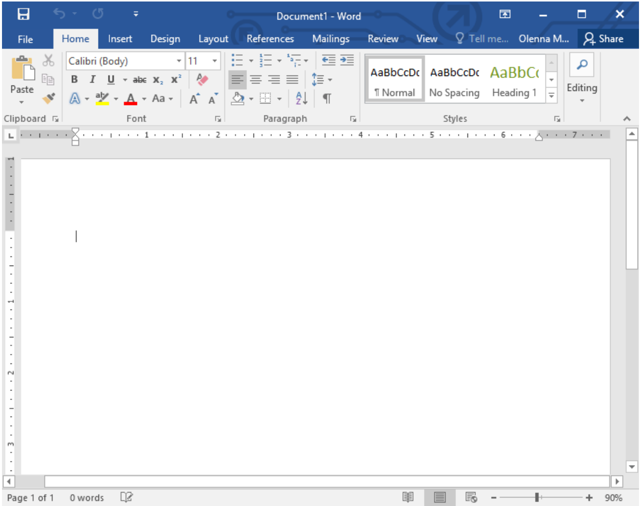


Figure 1: Microsoft Word Home Screen

Alternatives to Microsoft Word

1. **Google Docs** – Free, browser-based, real-time collaboration.
2. **LibreOffice Writer** – Free, offline, open-source, Word-compatible.
3. **WPS Writer** – Free version available, Excel-like interface.
4. **Overleaf (LaTeX editor)** – Best for academic research writing with formulas and structured formatting.

Spreadsheets

- A **spreadsheet** is a software application designed to organize, calculate, and analyze data in tabular form.
- Data is entered into a **grid of rows and columns**, forming **cells**.
- Each cell can contain text, numbers, or formulas.
- Spreadsheets are particularly useful for:
  - **Numerical analysis** (budgets, statistical summaries).
  - **Data visualization** (charts and graphs).
  - **Data management** (sorting, filtering, and summarizing).
- Popular spreadsheets include **Microsoft Excel, Google Sheets, and LibreOffice Calc**.

Introduction to Excel

- Microsoft Excel is a **spreadsheet program** used to store, organize, and analyze data.
- Data is arranged in **rows (numbers)** and **columns (letters)** forming **cells**.
- Each cell can contain:
  - **Text** (names, labels)
  - **Numbers** (data values)
  - **Formulas** (calculations)

Introduction to Excel

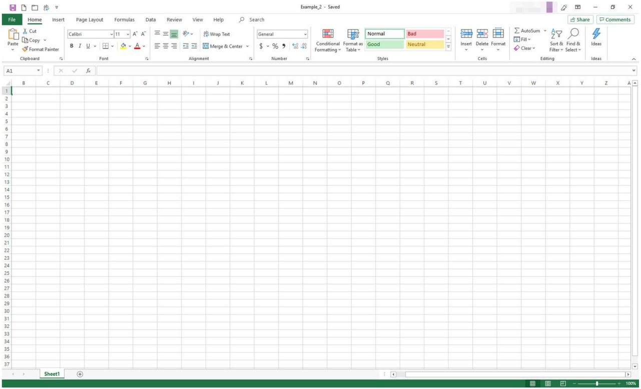


Figure 2: An Excel sheet

Excel Interface

- **Workbook** → The whole Excel file.
- **Worksheet** → A single tab/page inside a workbook.
- **Cell** → Intersection of a row and a column.
- **Cell Reference**:
  - `A1` = Column A, Row 1
  - `B5` = Column B, Row 5

Entering Data

- Click on a cell and type a value or text.
- Press **Enter** to go down, **Tab** to move right.
- Data types:
  - **Numeric**: 120, 3.75
  - **Text**: “Dhaka”, “Student”
  - **Date/Time**: 12/09/2025, 10:30 AM

Basic Formulas

- Always start with **=**.
- Examples:
  - `=A1 + B1` → Adds two cells.
  - `=A1 * B1` → Multiplies values.
  - `=A1 - B1` → Subtracts values.
  - `=A1 / B1` → Divides values.

Common Functions

- **SUM** → `=SUM(A1:A5)` adds all numbers from A1 to A5.
- **AVERAGE** → `=AVERAGE(B1:B10)` finds mean.
- **MAX / MIN** → `=MAX(C1:C20)`, `=MIN(C1:C20)` finds maximum and minimum.
- **COUNT** → `=COUNT(D1:D50)` counts numeric entries.

Formatting Data

- Change **font, size, and color**.
- Use **bold/italic/underline** for emphasis.
- Align text left, right, or center.
- Format numbers as:
  - Currency
  - Percentage
  - Date

Charts in Excel

- Select data → Insert → Choose chart type.
- Common charts:
  - **Column/Bar chart –** compare categories.
  - **Pie chart –** show proportions.
  - **Line chart –** show trends over time.

Example Exercise

- **Q1:** Enter 5 students' marks in Excel and calculate:
  - Total marks using `SUM()`.
  - Average marks using `AVERAGE()`.
  - Highest mark using `MAX()`.
- **Q2:** Create a bar chart of the marks.

Alternatives to Excel

1. **Google Sheets** – Free, online, real-time collaboration.
2. **LibreOffice Calc** – Free, offline, Excel-compatible.
3. **WPS Spreadsheets** – Free, Excel-like user interface.
4. **Zoho Sheet** – Cloud-based, business-oriented, integrates with Zoho apps.

Subsection 2

1.2 Statistical Software Overview

Introduction

- In modern research, as **data analysis becomes larger and more complex**, specialized **statistical software** has become essential.
- These tools help researchers and students:
  - Manage and clean large datasets efficiently.
  - Perform statistical tests, modeling, and predictive analysis.
  - Create professional-quality graphs and formatted reports.
- Each software package has **its own strengths and weaknesses**, and the choice often depends on:
  - Field of study,
  - Research objectives,
  - Budget,
  - Familiarity with coding.

Introduction

Figure 3: Different Statistical Software

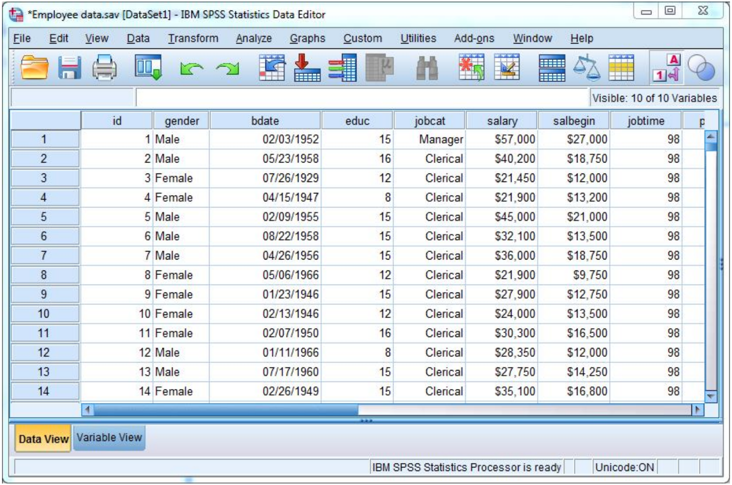
Major Categories of Statistical Software

1. **Menu-driven software** (ideal for beginners, point-and-click interfaces).
    - Examples: **SPSS, Minitab, JMP**.
    - Good for learning basics and running standard analyses.
2. **Syntax/programming-based software** (flexible and powerful, requires coding).
    - Examples: **R, Python, Stata, SAS**.
    - Preferred in advanced research and reproducible workflows.
3. **Hybrid software** (combines menus with scripting options).
    - Example: **SPSS** (menus + syntax).
    - Lets users begin with menus and gradually move to coding.

IBM SPSS Statistics

- **Strengths:**
  - Very user-friendly with a **point-and-click interface**.
  - Popular in **social sciences, psychology, health, and education**.
  - Strong for: data entry, descriptive statistics, regression, factor analysis.
  - Integrates easily with Excel spreadsheets.
- **Limitations:**
  - Less flexible for advanced/custom analyses.
  - Expensive (licensed software).
  - Slower when handling very large datasets.

IBM SPSS Statistics



*Employee data.sav [DataSet1] - IBM SPSS Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Custom
Utilities
Add-ons
Window
Help
Visible: 10 of 10 Variables
Data View
Variable View
IBM SPSS Statistics Processor is ready
Unicode:ON

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	p
1	1	Male	02/03/1952	15	Manager	$57,000	$27,000	98	
2	2	Male	05/23/1958	16	Clerical	$40,200	$18,750	98	
3	3	Female	07/26/1929	12	Clerical	$21,450	$12,000	98	
4	4	Female	04/15/1947	8	Clerical	$21,900	$13,200	98	
5	5	Male	02/09/1955	15	Clerical	$45,000	$21,000	98	
6	6	Male	08/22/1958	15	Clerical	$32,100	$13,500	98	
7	7	Male	04/26/1956	15	Clerical	$36,000	$18,750	98	
8	8	Female	05/06/1966	12	Clerical	$21,900	$9,750	98	
9	9	Female	01/23/1946	15	Clerical	$27,900	$12,750	98	
10	10	Female	02/13/1946	12	Clerical	$24,000	$13,500	98	
11	11	Female	02/07/1950	16	Clerical	$30,300	$16,500	98	
12	12	Male	01/11/1966	8	Clerical	$28,350	$12,000	98	
13	13	Male	07/17/1960	15	Clerical	$27,750	$14,250	98	
14	14	Female	02/26/1949	15	Clerical	$35,100	$16,800	98	

Figure 4: SPSS Data View

R (Free & Open Source)

- **Strengths:**
  - Free and open source (no license fees).
  - Extremely flexible, supports almost every statistical method.
  - Excellent for **data visualization** (e.g., *ggplot2*), interactive apps (*Shiny* dashboards).
  - Huge ecosystem of packages (CRAN, Bioconductor).
- **Limitations:**
  - Requires programming knowledge (steep learning curve).
  - Less intuitive for complete beginners.
  - Needs installation and management of packages.

Example: R is ideal for an academic researcher who needs advanced modeling (like survival analysis or machine learning) and is comfortable with coding.

Python (General-Purpose + Data Science)

- **Strengths:**
  - Free and open source.
  - Great for **data science, machine learning, and automation**.
  - Libraries like **pandas, NumPy, scikit-learn, TensorFlow, PyTorch** make it powerful.
  - Strong integration with databases, websites, and cloud platforms.
  - Popular in both academia and industry.
- **Limitations:**
  - Requires programming skills.
  - Fewer ready-made built-in statistical procedures compared to R.
  - Visualization requires additional libraries (*matplotlib, seaborn*).

Example: Python is widely used by tech companies for tasks like building recommendation systems, predictive modeling, and big data pipelines.

Stata

- **Strengths:**
  - Popular in **economics, epidemiology, sociology, and political science**.
  - Balanced between menus and scripting (easy to learn commands).
  - Excellent for **panel data analysis, survival models, and econometrics**.
  - Strong official documentation and user community.
- **Limitations:**
  - Commercial (paid) software.
  - Graphics less advanced compared to R/Python.
  - User community smaller than R/Python.

Example: An economist analyzing household panel data (like income and expenditure surveys) would find Stata very effective.

SAS

- **Strengths:**
  - Industry standard in **clinical trials, banking, insurance, and government**.
  - Extremely powerful for **big datasets** and **regulatory reporting**.
  - Offers advanced modeling, business analytics, and automation.
- **Limitations:**
  - Very expensive licensing costs.
  - Complex and less intuitive for beginners.
  - Declining popularity among students (due to free alternatives).

Example: A pharmaceutical company analyzing drug trial data for FDA submission often uses SAS due to its reliability and industry acceptance.

Comparison of Statistical Software

Software	Cost	Ease of Use	Best For	Limitations
SPSS	Paid	Very Easy	Social sciences, health	Expensive, limited flexibility
R	Free	Hard	Advanced stats, research	Steep learning curve
Python	Free	Medium	Data science, ML, automation	Needs programming
Stata	Paid	Medium	Economics, panel data	Cost, smaller community
SAS	Paid (very)	Hard	Pharma, industry, big data	Very expensive, complex

Practical Guidelines for Choosing Software

- If you are a beginner/social science student: Use **SPSS** (easy, reliable).
- If you want free and powerful tools for research: Choose **R or Python**.
- If working in economics, demography, or epidemiology: Choose **Stata**.
- If working in pharma or large-scale industry projects: Choose **SAS**.

Example Exercise 1

Question: A psychology researcher wants to do descriptive statistics and regression analysis without writing code. Which software would you recommend? Why?

Answer: SPSS – It has an easy point-and-click interface, is widely used in psychology, and requires no coding.

Example Exercise 2

Question: Match the field with the most common software:

1. Clinical trials in the pharmaceutical industry → ?
2. Machine learning models → ?
3. Teaching basic statistics in undergraduate classes → ?

Answer:

1. SAS.

2. Python (or R).

3. SPSS (or Minitab).

Summary

- A variety of **statistical software packages** exist: SPSS, R, Python, Stata, SAS, Minitab, JMP.
- Each has **specific strengths and limitations** depending on cost, ease of use, and intended purpose.
- **SPSS & Minitab** → Beginner-friendly, ideal for teaching and social sciences.
- **R & Python** → Free, flexible, excellent for research and data science, but require coding.
- **Stata & SAS** → Specialized for certain research and industry fields.
- Choosing the right software depends on **your research field, budget, and level of technical expertise**.

Subsection 3

1.3 SPSS Background

Introduction to SPSS

- **SPSS (Statistical Package for the Social Sciences)** is one of the most widely used statistical software programs.
- Originally developed in the late 1960s, it is now owned by **IBM** and officially called **IBM SPSS Statistics**.
- Commonly used in **social sciences, psychology, health, education, business, and market research**.
- Provides two modes of working:
  - **Menu-driven interface (point-and-click)** – easy for beginners.
  - **Syntax (command language)** – ensures reproducibility for advanced users.
- SPSS include **data visualization, advanced statistical tests, predictive models, and reporting tools**.

Why SPSS Became Popular

- **Ease of Use**: Point-and-click interface makes it accessible to beginners without coding.
- **Reproducibility**: Syntax editor allows advanced users to document and repeat analyses.
- **Versatility**: Handles descriptive statistics, hypothesis testing, regression, multivariate methods, and time-series analysis.
- **Integration**: Can import/export data from Excel, CSV, Stata, SAS, and other formats.
- **Professional Output**: Produces clean, well-formatted tables and charts ready for reports or publications.
- **Wide Acceptance**: Adopted by universities, NGOs, and government agencies worldwide, especially in survey and behavioral research.
- **Consistency and Reliability**: Established a reputation for stable, trusted results, making it a standard in academic and applied fields.

Example: A public health researcher can quickly import survey data, run chi-square tests, and generate graphs for a report, all without programming, demonstrating why SPSS became a preferred tool.

Applications of SPSS

1. **Data Management**
   - Data entry and cleaning.
   - Handling missing values.
   - Recoding and computing new variables.
2. **Descriptive Statistics**
   - Frequency tables and cross-tabulations.
   - Mean, median, mode, variance, standard deviation.
3. **Inferential Statistics**
   - Hypothesis testing (*t-test, chi-square, ANOVA*).
   - Correlation and regression.
   - Logistic regression and non-parametric tests.

Applications of SPSS

4. **Advanced Analysis**
   - Factor analysis, PCA, and reliability analysis.
   - Multivariate methods (MANOVA, discriminant analysis).
   - Time-series forecasting (ARIMA, exponential smoothing).
5. **Visualization**
   - Charts and graphs (bar charts, histograms, scatter plots).
   - Boxplots and cluster plots.
   - Pivot tables for summaries.

Strengths of SPSS

- Beginner-friendly.
- Produces professional, publication-ready outputs.
- Strong in survey-based and questionnaire research.
- Well-documented with training resources.
- Trusted in both academia and industry.

Limitations of SPSS

- Paid software, relatively expensive.
- Less flexible compared to open-source tools like **R** or **Python**.
- Can be slow with very large datasets.
- Limited in machine learning and AI applications.

For modern predictive modeling, R or Python may be better options, but SPSS remains excellent for classic statistical analysis.

Example Exercise

Question: A researcher has survey data from 200 students on study habits and exam scores. Suggest three analyses they could do in SPSS.

Answer:

1. Descriptive statistics of study hours (mean, SD).
2. Cross-tabulation of gender × study habits.
3. Linear regression predicting exam score from study hours.

Summary

- **SPSS** is a long-established, reliable, and user-friendly statistical software.
- Best for **survey analysis, descriptive and inferential statistics, and basic modeling**.
- GUI makes it accessible for beginners, while syntax helps advanced users.
- Despite limits in machine learning, SPSS continues to be a **cornerstone of applied research and teaching** worldwide.

Subsection 4

1.4 SPSS Windows

Starting SPSS Statistics

At first, download and install the SPSS installer from the IBM SPSS website (license purchase required).

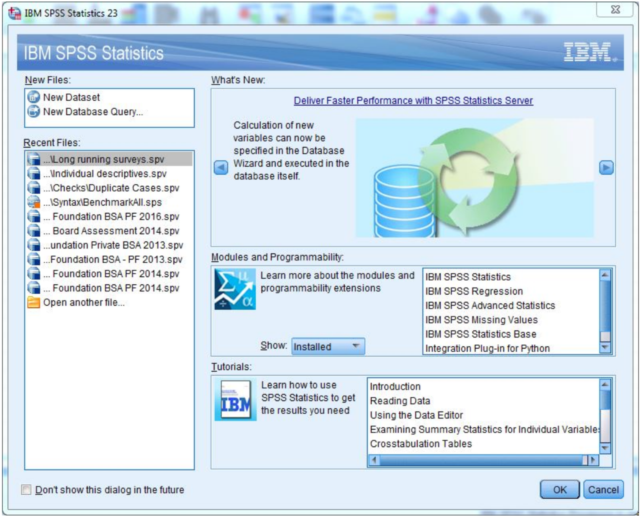
To launch **SPSS Statistics** on a Windows computer:

`Start Menu > All Programs > IBM SPSS Statistics > IBM SPSS Statistics 31`

When SPSS starts for the first time, you will see an **initial dialog box**. This dialog asks you whether you want to:

- Open a recently used file, or
- Open another file from your computer, or
- Create a new file from scratch.

Starting SPSS Statistics



IBM SPSS Statistics 23

IBM SPSS Statistics

New Files:

- New Dataset
- New Database Query...

Recent Files:

- ...\Long running surveys.spv
- ...\Individual descriptives.spv
- ...\Checks\Duplicate Cases.spv
- ...\Syntax\BenchmarkAll.sps
- ... Foundation BSA PF 2016.spv
- ... Board Assessment 2014.spv
- ...undation Private BSA 2013.spv
- ...Foundation BSA - PF 2013.spv
- ... Foundation BSA PF 2014.spv
- ... Foundation BSA PF 2014.spv
- Open another file...

What's New:

Deliver Faster Performance with SPSS Statistics Server

Calculation of new variables can now be specified in the Database Wizard and executed in the database itself.

Modules and Programmability:

Learn more about the modules and programmability extensions

Show: Installed

- IBM SPSS Statistics
- IBM SPSS Regression
- IBM SPSS Advanced Statistics
- IBM SPSS Missing Values
- IBM SPSS Statistics Base
- Integration Plug-in for Python

Tutorials:

Learn how to use SPSS Statistics to get the results you need

- Introduction
- Reading Data
- Using the Data Editor
- Examining Summary Statistics for Individual Variable:
- Crosstabulation Tables

□ Don't show this dialog in the future

In most cases, you will begin your SPSS session by opening the **data file** you want to work with (for example, survey data, experimental data, or secondary datasets).

Three Main Windows in SPSS

SPSS Statistics works through **three main windows**, plus a **menu bar** at the top.

These windows allow you to:

1. Enter and view your data,
2. View the statistical results,
3. Write or run commands.

Each of these windows is linked to a different SPSS file type.

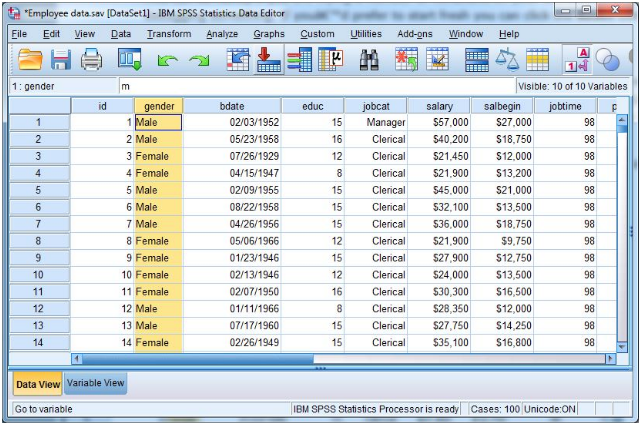
1. Data Editor Window (.sav files)

The **Data Editor** is where you enter, edit, and view your dataset.

The Data Editor gives you two views of your data set: a *Data View* and a *Variable View*, selected by clicking on the appropriate tab in the lower left corner of the window.

1. **Data View**
   - Looks like a spreadsheet.
   - **Rows** = individual cases (e.g., each respondent, household, or patient).
   - **Columns** = variables (e.g., age, gender, income, education).
   - You can type directly into cells or paste data from Excel.

1. Data Editor Window (.sav files)



*Employee data.sav [DataSet1] - IBM SPSS Statistics Data Editor

1 : gender   m

Visible: 10 of 10 Variables

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	p
1	1	Male	02/03/1952	15	Manager	$57,000	$27,000	98	
2	2	Male	05/23/1958	16	Clerical	$40,200	$18,750	98	
3	3	Female	07/26/1929	12	Clerical	$21,450	$12,000	98	
4	4	Female	04/15/1947	8	Clerical	$21,900	$13,200	98	
5	5	Male	02/09/1955	15	Clerical	$45,000	$21,000	98	
6	6	Male	08/22/1958	15	Clerical	$32,100	$13,500	98	
7	7	Male	04/26/1956	15	Clerical	$36,000	$18,750	98	
8	8	Female	05/06/1966	12	Clerical	$21,900	$9,750	98	
9	9	Female	01/23/1946	15	Clerical	$27,900	$12,750	98	
10	10	Female	02/13/1946	12	Clerical	$24,000	$13,500	98	
11	11	Female	02/07/1950	16	Clerical	$30,300	$16,500	98	
12	12	Male	01/11/1966	8	Clerical	$28,350	$12,000	98	
13	13	Male	07/17/1960	15	Clerical	$27,750	$14,250	98	
14	14	Female	02/26/1949	15	Clerical	$35,100	$16,800	98	

Go to variable

IBM SPSS Statistics Processor is ready

Cases: 100

Unicode:ON

Figure 5: SPSS Data View

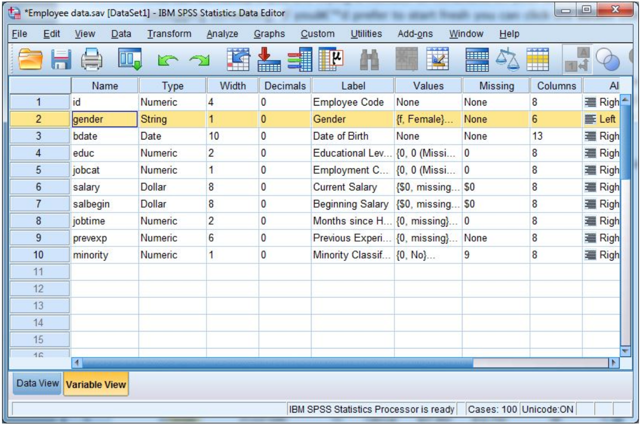
Example: If you are analyzing 100 students' exam marks, each row represents one student, and each column represents a variable (such as `ID`, `Gender`, `Exam_Score`).

1. Data Editor Window (.sav files)

2. **Variable View**
   - Used to **define and manage variables**.
   - Columns in Variable View include:

     Name: short name of the variable (no spaces allowed).
     Type: numeric, string, date, etc.
     Label: descriptive label for the variable.
     Values: codes and labels (e.g., 1 = Male, 2 = Female).
     Measure: nominal, ordinal, or scale.
   - Important for ensuring your data is analyzed correctly.

1. Data Editor Window (.sav files)



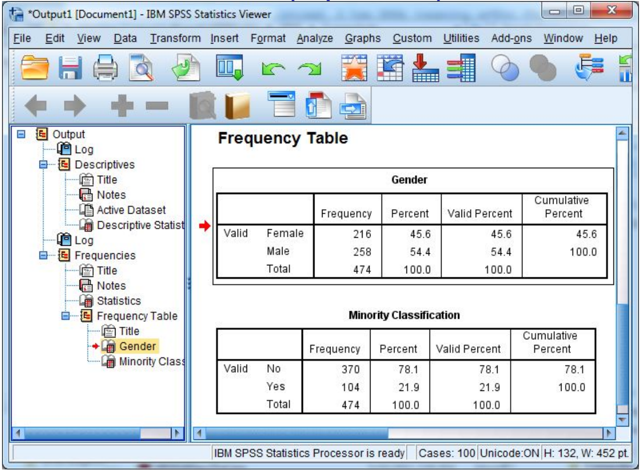
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Al
1	id	Numeric	4	0	Employee Code	None	None	8	Righ
2	gender	String	1	0	Gender	{f, Female}...	None	6	Left
3	bdate	Date	10	0	Date of Birth	None	None	13	Righ
4	educ	Numeric	2	0	Educational Lev...	{0, 0 (Missi...	0	8	Righ
5	jobcat	Numeric	1	0	Employment C...	{0, 0 (Missi...	0	8	Righ
6	salary	Dollar	8	0	Current Salary	{$0, missing...	$0	8	Righ
7	salbegin	Dollar	8	0	Beginning Salary	{$0, missing...	$0	8	Righ
8	jobtime	Numeric	2	0	Months since H...	{0, missing}...	0	8	Righ
9	prevexp	Numeric	6	0	Previous Experi...	{0, missing}...	None	8	Righ
10	minority	Numeric	1	0	Minority Classif...	{0, No}...	9	8	Righ
11									
12									
13									
14									
15									

Figure 6: SPSS Variable View

2. Output Viewer Window (.spv files)

- This is where the results of your analysis appear.
- It includes **tables, charts, and statistical test results**.
- You can **copy** results into Word or PowerPoint for reports, or **export** them to PDF, Excel, or HTML.

2. Output Viewer Window (.spv files)


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Custom Utilities Add-ons Window Help
Output
Log
Descriptives
Title
Notes
Active Dataset
Descriptive Statist
Log
Frequencies
Title
Notes
Statistics
Frequency Table
Title
Gender
Minority Class

IBM SPSS Statistics Processor is ready | Cases: 100 | Unicode:ON | H: 132, W: 452 pt.

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	

Minority Classification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	370	78.1	78.1	78.1
	Yes	104	21.9	21.9	100.0
	Total	474	100.0	100.0	

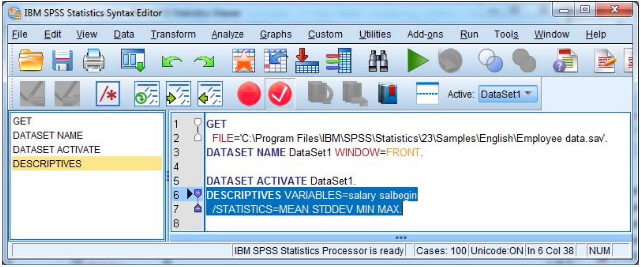
Figure 7: SPSS Output Viewer with frequency table

Example: After running a frequency analysis on the variable “Gender”, a table will appear in the Output Viewer showing the number and percentage of males and females.

3. Syntax Editor Window (.sps files)

- The **Syntax Editor** is for writing commands in SPSS language.
- Commands are saved in files ending with `.sps`.
- Useful for repeating analyses without clicking menus again and again.
- Encouraged for advanced users because it ensures reproducibility.

3. Syntax Editor Window (.sps files)



```
GET
  FILE='C:\Program Files\IBM\SPSS\Statistics\23\Samples\English\Employee data.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.

DATASET ACTIVATE DataSet1.
DESCRIPTIVES VARIABLES=salary salbegin
  /STATISTICS=MEAN STDDEV MIN MAX.
```

IBM SPSS Statistics Processor is ready | Cases: 100 | Unicode:ON | In 6 Col 38 | NUM

Example command:

```
FREQUENCIES VARIABLES=Gender.
```

This generates the same frequency table as the point-and-click method.

Other Windows

- **Chart Editor**: lets you modify graphs (e.g., change colors, add titles, edit axes).
- **Pivot Table Editor**: allows you to reformat output tables (change fonts, merge cells, add totals).

These editors are powerful tools for preparing results for reports or publications.

Advantages of SPSS's Window System

- Clear separation of **data, metadata, and results** helps you stay organized.
- Very **beginner-friendly** because of its graphical interface.
- Offers **flexibility**:
  - *Point-and-click menus* for beginners,
  - *Syntax commands* for advanced users and reproducibility.

File Types in SPSS

Extension	File Type	Contents
`.sav`	Data File	Contains data + variable definitions
`.sps`	Syntax File	Stores analysis commands/scripts
`.spv`	Output File	Stores results of statistical analyses

Note: `.sav` files are the most commonly used. If you only save your output (.spv), you won't be able to re-run your analysis later without the data file.

Example Exercise

Question: Match each SPSS window with its main purpose:

1. Data entry and editing → ?
2. Variable definition → ?
3. Results display → ?

Answer:

1. Data Editor: Data View.
2. Data Editor: Variable View.
3. Output Viewer.

Summary

- The **three main windows** are:
  1. **Data Editor** (Data View + Variable View),
  2. **Output Viewer**,
  3. **Syntax Editor**.
- Additional editors help with customizing charts and tables.
- The separation of data, syntax, and output makes SPSS **easy to learn and reliable for research**.