STATISTICAL ANALYSIS USING SPSS

SPSS stands for Statistical Package for the Social Sciences. This program can be used to analyze data collected from surveys, tests, observations, etc. It can perform a variety of data analyses and presentation functions, including statistical analysis and graphical presentation of data. Among its features are modules for statistical data analysis. These include:

(1) descriptive statistics such as frequencies, central tendency, plots, charts, and lists; and

(2) sophisticated inferential and multivariate statistical procedures such as analysis of variance (ANOVA), factor analysis, cluster analysis, and categorical data analysis.

Overview of the User Interface

The Data Editor window opens with two view tabs: Data View and Variable View. Data View is used for data input, and Variable View is used for adding variables and defining variable properties (e.g., modifying attributes of variables). As displayed in Figure 1, the Data Editor window includes several components. The Title bar displays the name of the current file and the application. The Menubar provides access to various commands which are grouped according to function. The Data Editor toolbar provides shortcuts to commonly used menu commands.
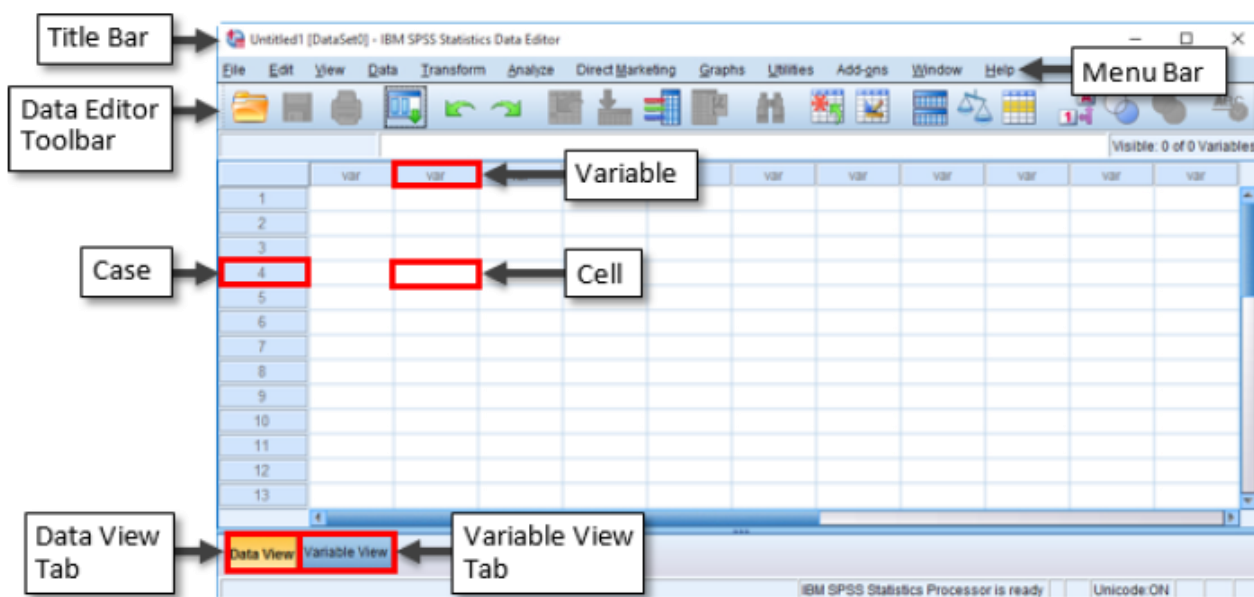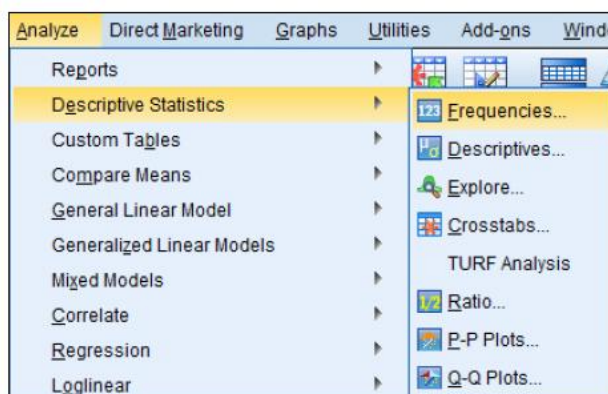

Figure 3 – IBM SPSS Statistics Data Editor Window

Descriptive Statistics

After data has been entered, it can be analyzed using descriptive statistics. Descriptive statistics is commonly used for summarizing data frequency or measures of central tendency (mean, median, and mode)
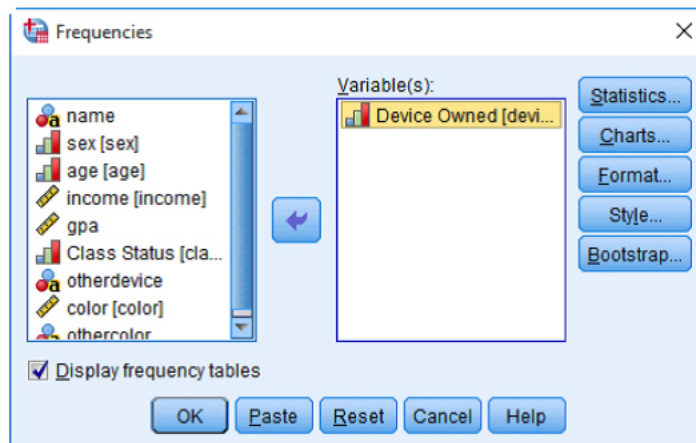
Steps to calculate the descriptive statistics:

1. Click the **Open** button on the **Data Editor** toolbar.

2. In the **Open Data** dialog box, navigate to the location where you saved the data files, select the file, and then click the **Open** button.

3. Click the **Analyze** menu, point to **Descriptive Statistics**, and then click **Frequencies**.

Figure– Frequencies Selected on the Analyze Menu

4. In the **Frequencies** dialog box, select the variable(s) that you want to analyze. In this case, select the **Device Owned** variable in the box on the left, and then click the transfer arrow button. The selected variable is moved to the **Variable(s)** box.

5. Make sure that the **Display frequency tables** check box is selected.
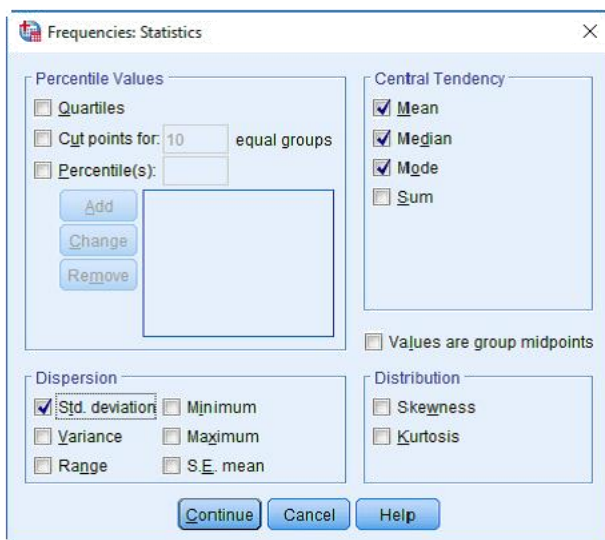
6. Click the **Statistics** button.

Frequencies Dialog Box

7. In the **Frequencies: Statistics** dialog box, in the **Central Tendency** section, select the **Mean**, **Median**, and **Mode** check boxes.

8. In the **Dispersion** section, select the **Std. deviation** check box.

9. Click the **Continue** button.

10. In the **Frequencies** dialog box, click the **OK** button. The **Output Viewer** window opens and displays the statistics and frequency tables. The columns of the **Device Owned** table display the **Frequency**, **Percent**, **Valid Percent**, and **Cumulative Percent** for each different type of device owned

Output will be obtained as follows:

Statistics

Device Owned

N	Valid	74
	Missing	6
Mean		3.07
Median		3.00
Mode		3
Std. Deviation		.912

Device Owned

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Tablet	3	3.8	4.1	4.1
	Mac	10	12.5	13.5	17.6
	PC	49	61.3	66.2	83.8
	Smartphone	3	3.8	4.1	87.8
	Other	9	11.3	12.2	100.0
	Total	74	92.5	100.0	
Missing	System	6	7.5		
Total		80	100.0		

The measures of central tendency (mean, median, and mode) can be used to summarize various types of data. Mode can be used for nominal data such as device type, device color, ethnicity, etc. Mean or median can be used for interval/ratio data such as test scores, age, etc. The mean is also useful for data with a skewed distribution.

Advantages: The **advantages** of using SPSS for statistical analysis are:

• SPSS is a comprehensive statistical software which can be used for both simple and complex analysis.

• Many statistical tests are available as a built in feature in the program which suits the needs of the research communities.

• Interpretation of results is relatively easy.

• It easily and quickly displays data tables.

• It can be expanded according to the need of the statistical analysis.
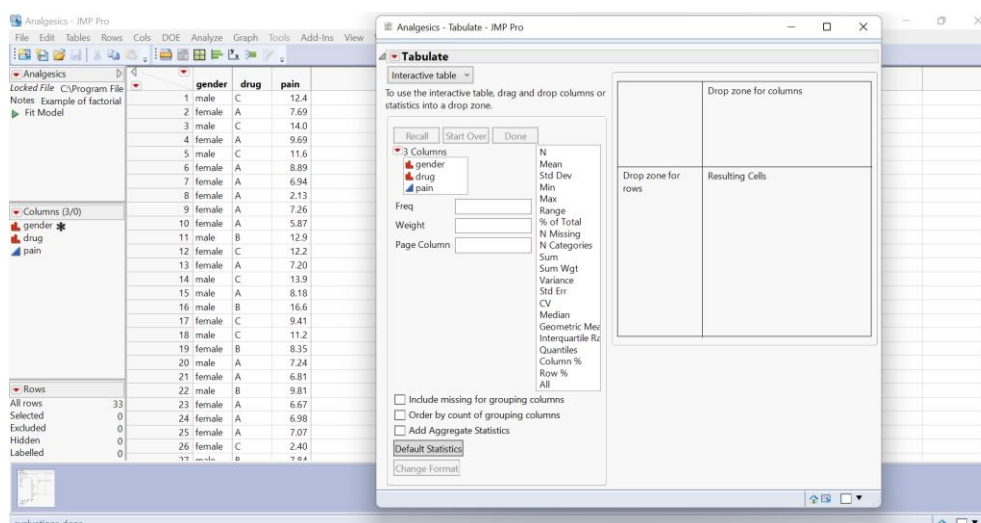
LIMITATIONS

• SPSS is a paid software and can be expensive for students requiring limited use.

• Usually involves added training to completely exploit all the available features.

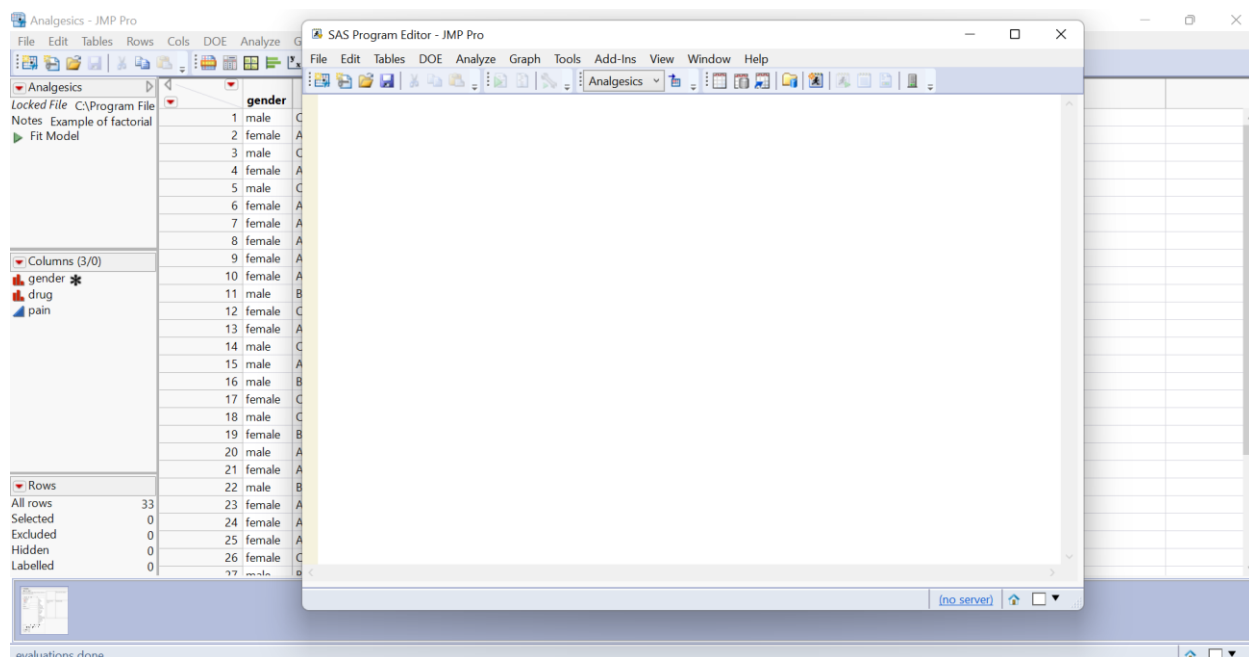• The graph features are not as simple as that of Microsoft Excel

Introduction to SAS-JMP:

SAS-JMP is a go-to application for researchers, data analysts, engineers, and scientists in a wide range of industries. Its data analysis capabilities allow for a comprehensive treatment of complex data that is not visually dull. JMP has an entire suite for statistical discovery. It allows users to notice patterns and extract insights. Complete with data visualization options, reports can be generated easier for faster and more accurate response times. It is being used by many types of professionals today in growing their financial assets, monitoring species, and the creation of greener products. JMP makes doing all of these easier by allowing users to access data from different sources and keeping everything organized in one hub. The quick and accurate data preparation modules also help them save time and conduct their analyses better. Complex relationships can then be mapped to different variables to help users determine existing and possible relationship schemes. Furthermore, these can be easily shared with others. JMP supports web visualizations and comes with an interactive dashboard. Predictive models are very important in virtually all types of industries, enterprises, and organizations. JMP makes predictive analysis easier and more accurate in the sense that it offers users tools to acquire more data-driven actionable insights. There is also a feature where clinical analyses and reports can be streamlined. JMP has a wide range of tools that accommodate organizations and projects of all types from business intelligence to the handling of genomic data.

Interface:

Main screen interface

SAS script editor

Features:

Main features of SAS-JMP are:

- Advanced Computational Statistics
- Automation & Scripting
- Biometrics & Biostatistics
- Clinical Data Analysis
- Consumer & Market Research
- Dat Acquisition
- Data Management & Integrity
- Data Monitoring
- Data Visualization
- Design Of Experiments
- Expression
- Formula Depot & Score Code
- Genomic Data Analysis
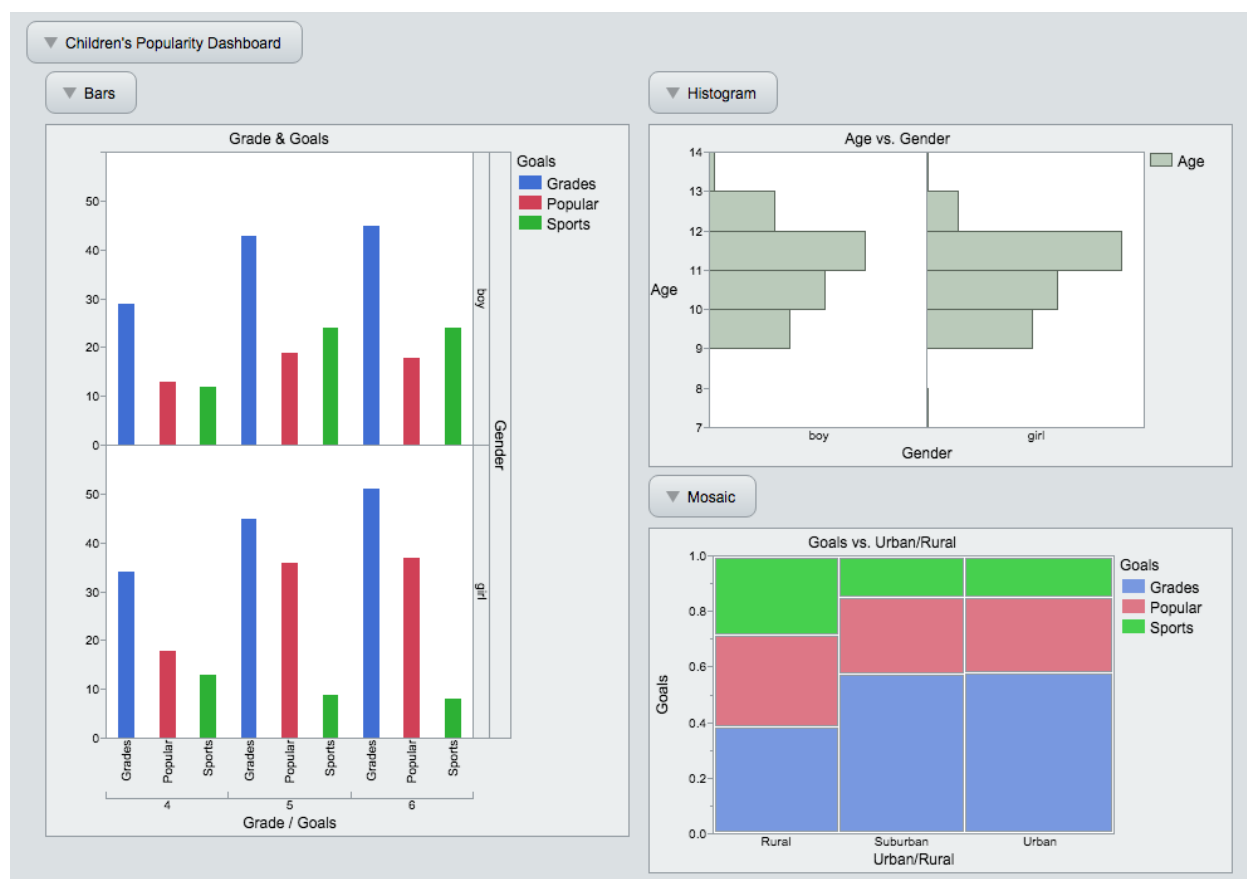- Genomic Selection
- Crop Improvement

- Group, Filter, And Subset Data
- Linkage Mapping
- Mixed Models
- Model Comparison
- Next-Gen Sequencing
- Patient Profiles & Narratives
- Pharmacogenomics
- Predictive Modeling
- Cross-Validation
- Quality & Process Engineering
- Repairable Systems Simulation
- Risk-Based Monitoring
- Share & Communicate Results
- Statistical Genetics
- Text Exploration

Benefits and advantages:

The main benefits of JMP are its data analysis features, data visualization tools, and business intelligence.

- Users get to build models to help them understand their data better through JMP's predictive modeling tools. Trends, expectations, priors, and weights can be assigned to simulate time sequences to project possible scenarios. This has a wide range of applications and has been in demand in many types of ventures from agriculture to the medical field.

- There is also a version of JMP that help users simplify processes of drug development. Review and reporting of clinical trials data can be automated. With summary dashboards, users can easily monitor data and evaluate them. This also helps medical writers in being able to quickly produce patient profiles and narratives for review by internal panels and even corresponding regulatory agencies.

- Scientists in the life sciences also benefit from JMP's advanced genomics data analysis capabilities. Those who work on crops and plant breeds are able to discover new combinations and trait variations. This can help them in the design of more resilient strain of crops. Furthermore, this tool can be used by scientists working on the genetic patterns and underpinnings of drug responses and diseases. There are numerous ways to deploy this particular set of tools for different projects and JMP supports them all.

- JMP also helps users in cross-validation where statistical models are refined in ways that they can be used more generally for other samples. The software helps users avoid overfitting and properly differentiate actual patterns from the noise.

- Data analysis would almost amount to nothing if it is not properly communicated to other researchers or management. Data-driven reports can now be easily shared using JMP in digestible formats. Trends and patterns can be reported with the aid of data visualization. Furthermore, this can be in a dynamic and interactive way. With the graph builder, visualizations can be manipulated in ways that can help users understand the complex relationships and processes that they represent easier. With the Bubble Plot tool, users can animate data and present them as movies where changes in dimensions over time can be highlighted. With JMP, stories behind the data can be shared more easily.

Limitations:

- Learning curve is long and requires training for understanding all the features

- Expensive software

- Cross-platform compatibility is poor

- Not user-friendly

Epi-info:

Epi Info is a public domain suite of interoperable software tools for Microsoft Windows, designed for the global community of public health practitioners and researchers. It provides easy data entry form and database construction, a customized data entry experience, and data analyses with epidemiologic statistics, maps, and graphs for public health professionals who may lack an informational technology background.

Epi Info is used for outbreak investigation, for developing small to mid-sized disease surveillance systems, as analysis, visualization and reporting (AVR) components of larger systems, and in the continuing education in the science of epidemiology and public health analytic methods at schools of public health around the world.

Originally Epi Info was developed for the Centers for Disease Control and Prevention (CDC) own investigators. It was meant to provide tools for Epidemiologist and field investigators including, sample size calculators, 2X2 table calculators and questionnaires for case-control studies. Since 1998, there have been approximately 2,000,000 successful downloads of Epi Info from the CDC website in 180 countries. For many countries Epi Info provides the core surveillance system for imminizable disease for 27 African countries and covers over 25% of surveillance monitoring for Acute Flaccid Paralysis in India.[1]

Interface:

Features:

Form Designer – Form Designer and the Enter modules work together to design the data entry process and collect data. Form Designer is the tool used to design the survey, questionnaire, or form, tailor the data entry process, and specify the tab sequence. It is where you customize any data validation you want to occur when the form receives data in the Enter tool.

Enter Data – The Enter program displays Forms constructed in Form Designer.

Data Packager – The Epi Info Data Packager tool provides an easy way to share data with other users or to merge data collected by multiple users into a single database for analyses. The Data Packager does this by offering the option to package and export, as well as import data from Epi Info projects.

Compatibility with industry standards, including:

> Microsoft Access and other SQL and ODBC databases
> Visual Basic, Version 6
> Output as HTML web pages

Classic Analysis - Classic Analysis manipulates, manages, and analyzes data. It acts as a statistical toolbox providing many ways to transform data and perform statistical analyses. Data can be selected, sorted, listed, or manipulated with a series of commands, functions, and operators. Available statistics include frequencies, means, and more advanced processes (i.e., Kaplan-Meier Survival Classic Analysis and Logistic Regression).

Maps - Maps is a versatile tool that geographically displays data on a map. The Maps tool has the ability to display multiple views from the same dataset. Datasets can be filtered or shown over a series of time using features in the Maps tool. Users can tailor these features to create a customized map containing public health data.

Translation- Translation of Epi Info 7 programs, help files, and exercises into non-English languages is formulated by placing the translated phrase of all English phrases used in the programs in an MS Access.

Web Survey- Web Survey System allows the survey designer to collect information from participants over the internet.

Nutritional Anthropometry- Tools used to collect, analyze, and graph child growth data. Among these tools are a data entry form to calculate z-scores and percentiles as data are entered, growth charting capabilities in the Visual Dashboard, and several nutrition

functions in Classic Analysis that you can use to add z-scores and percentiles to existing sets of data.

Stat Calc- StatCalc is an epidemiologic calculator that produces statistics from summary data. Three types of calculations are offered:

Statistics from 2-by-2 to 2-by-9 tables similar to those produced in Analysis. Both single and stratified 2-by-2 tables can be analyzed to produce odds ratios and risk ratios (relative risks) with confidence limits, several types of chi square tests, Fisher exact tests, Mantel-Haenszel summary odds ratios and chi squares, and associated p-values.

Sample size and power calculations include Population Survey, Cohort or Cross-Sectional, and Unmatched Case-Control Study.

Chi-square for trend by the Mantel extension of the Mantel-Haenszel summary odds ratio and chi square. This tests for the presence of a trend in dose response or other case control studies where a series of increasing or decreasing exposures is being studied.

Advantages

1. Free for all users.
2. Easy to select subsets of data for analysis, without having to delete records or make multiple copies of datasets.
3. Keeps a saveable record of the analysis steps you have performed.
4. Automatically includes all VALID data points in calculations/tests.
5. Performs both descriptive statistics and a lot of basic to intermediate analysis (e.g., comparisonof means, proportions; many regression methods)
6. Can be used for data entry - uses an Access-like data storage system.

Disadvantages

1. Runs under Windows only.
2. Can be difficult to learn.
3. Limited analysis options beyond the basic methods.
4. Graphics can look quite sloppy and are difficult to scale/export for publication/presentation.