

IBM SPSS Statistics 23

Part 1: Descriptive Statistics

Winter 2016, Version 1

Table of Contents

Introduction	2
Downloading the Data Files.....	2
Starting the Program.....	2
Overview of the User Interface.....	3
Data View	4
Variable View.....	4
Creating a Data File.....	4
Defining Variables.....	5
Entering Data.....	7
Descriptive Statistics.....	7
Frequency Analysis.....	7
Crosstabs	10
Data Manipulation.....	12
Select Cases	12
Split File	15
Find and Replace	16
Reports.....	17
Appendix.....	18

Introduction

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. IBM SPSS Statistics 23 is well-suited for survey research, though by no means is it limited to just this topic of exploration.

This handout introduces the basic skills necessary to run IBM SPSS Statistics 23, including how to create a data file and run descriptive statistics. It gives users an overview of how SPSS Statistics can be used for survey research by answering three research questions formulated in the sample survey questionnaire.

The three research questions in the sample survey are:

1. What kind of device do people prefer to own?
2. What color do people prefer for their device?
3. Is device color preference different between males and females?

Downloading the Data Files

This handout includes sample data files that can be used to follow along the steps. If you plan to use the data files, download the following ZIP file to your computer and extract the files. It is recommended to save the data files on your desktop for easy access.

- [IBM SPSS Statistics 23 Part 1 Data Files](#)

Starting the Program

The following steps cover how to start IBM SPSS Statistics 23 from the Start menu in Windows 10.

To start IBM SPSS Statistics 23:

1. Click the **Start** button on the taskbar, click **All apps**, click **IBM SPSS Statistics**, and then click **IBM SPSS Statistics 23**.
2. If you are starting the program for the first time, the **Unicode Mode** dialog box opens. Click the **Use Unicode encoding** button (see Figure 1).

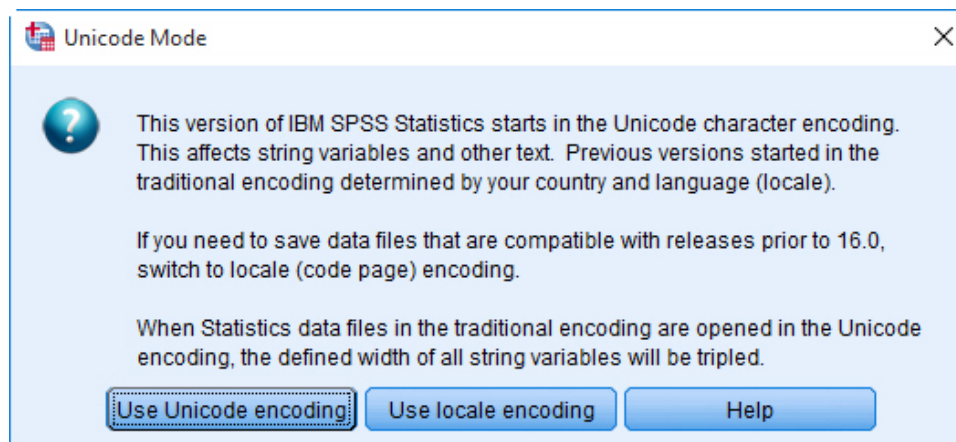


Figure 1 – Unicode Mode Dialog Box

- In the **IBM SPSS Statistics 23** dialog box, click the **Cancel** button (see Figure 2).

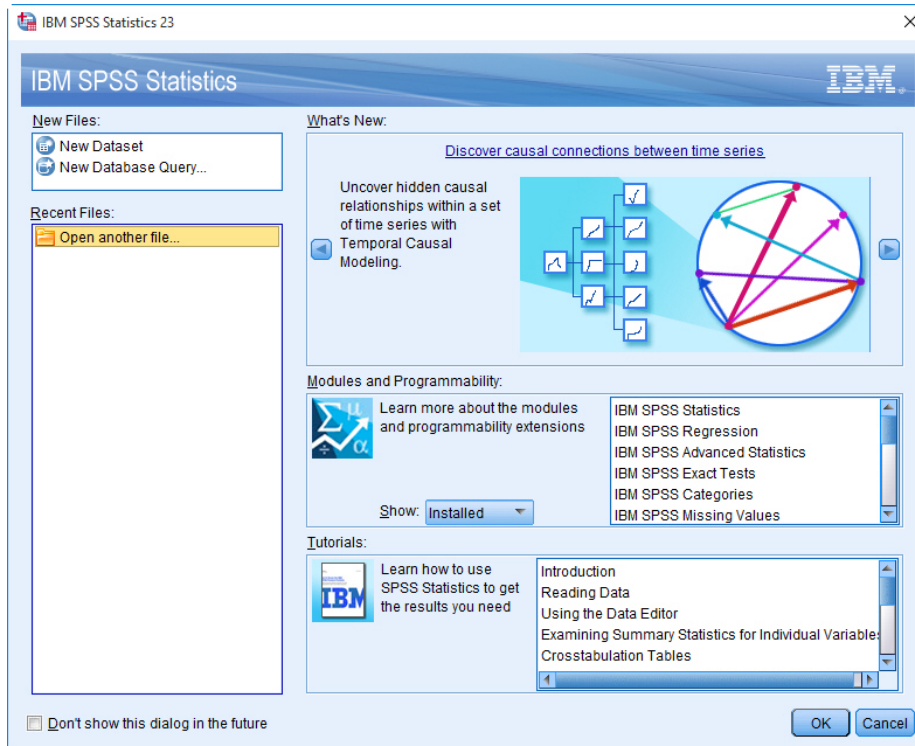


Figure 2 – IBM SPSS Statistics 23 Dialog Box

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 3, the Data Editor window includes several components. The **Title bar** displays the name of the current file and the application. The **Menu bar** provides access to various commands which are grouped according to function. The **Data Editor toolbar** provides shortcuts to commonly used menu commands.

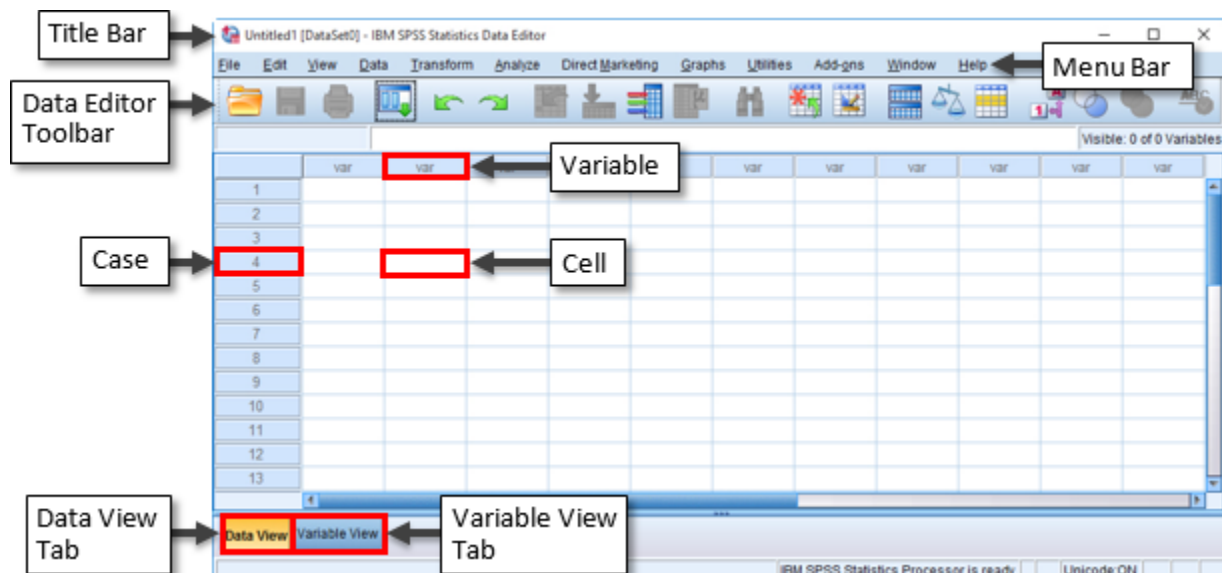


Figure 3 – IBM SPSS Statistics Data Editor Window

Data View

When SPSS Statistics is launched, the Data Editor window opens in **Data View** which looks similar to a Microsoft Excel worksheet (a matrix consisting of rows and columns). The difference is that the rows and columns in Data View are referred to as **cases** and **variables**, respectively (see Table 1).

Table 1 – Elements in Data View

Element	Description
Variable	Each column represents a variable. Any survey questionnaire item or test item can be a variable. Commonly defined variable types are numeric or string. When defining variables as numeric, users need to specify decimal places. Variable names can be up to 256 characters long and must start with a letter. Make variable names meaningful and easily recognizable.
Case	Each row represents a case. The participants in the study can be cases. For example, if 100 participants are involved in your study, then 100 cases (or rows) of information should be generated. Responses to the question items should be entered consistently from left to right for each participant.
Cell	A cell is an intersection between cases and variables. Each response to a survey question should be entered in a cell for each participant according to the defined variable data types.

Variable View

Variable View is where variables are defined by assigning variable names and specifying the attributes such as data type (**String**, **Date**, **Numeric**, etc.), value labels, and measurement scales (**Nominal**, **Ordinal**, or **Scale**). You can think of Variable View as the backbone structure for Data View; data cannot be entered nor viewed without first defining variables in Variable View (see Table 2).

Table 2 – Elements in Variable View

Element	Description
Variable name	SPSS Statistics initially assigns default variable names (VAR00001, VAR00002, etc.) which users can change. It is recommended to assign a brief and meaningful name to variables (e.g., Name, Sex, and GPA).
Variable type	The variable type determines how the cases are entered. Generally, text-based characters are of String type and number-based characters are of Numeric type. For example, if a user has a variable called Name, then its variable type should be String. Similarly, a variable named GPA should be a Numeric type, typically given two decimal places.
Value labels	Value labels allow users to describe what the variable name stands for. For example, a variable with an ambiguous name like Fav may confuse others working with the file. To avoid misinterpretation, value labels can be utilized to clearly define variable names.

Creating a Data File

Creating a new SPSS Statistics data file consists of two stages: (1) defining the variables and (2) entering the data. Defining the variables involves multiple processes and requires careful planning. Once the variables have been defined, the data can then be added.

Defining Variables

First, assign variable names based on your research questionnaire. If variable names are not assigned, SPSS Statistics provides default names that may not be recognizable. Second, each variable's **Type** attribute should be specified. If necessary, assign labels to values to help all users of the file better understand the data.

To define variables (example):

1. Click the **Variable View** tab in the lower-left corner of the **Data Editor** window (see Figure 4).

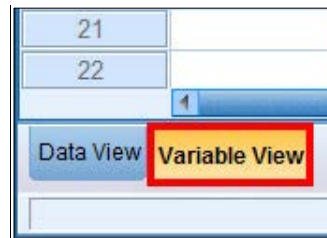



Figure 4 – Variable View Tab

2. Type **Name** in the first cell under the **Name** column, and then press the **Enter** key.
3. Under the **Type** column, click **Numeric**, and then click the **Ellipses** button  that appears in the cell.
4. In the **Variable Type** dialog box, select the **String** option button, and then click the **OK** button (see Figure 5).

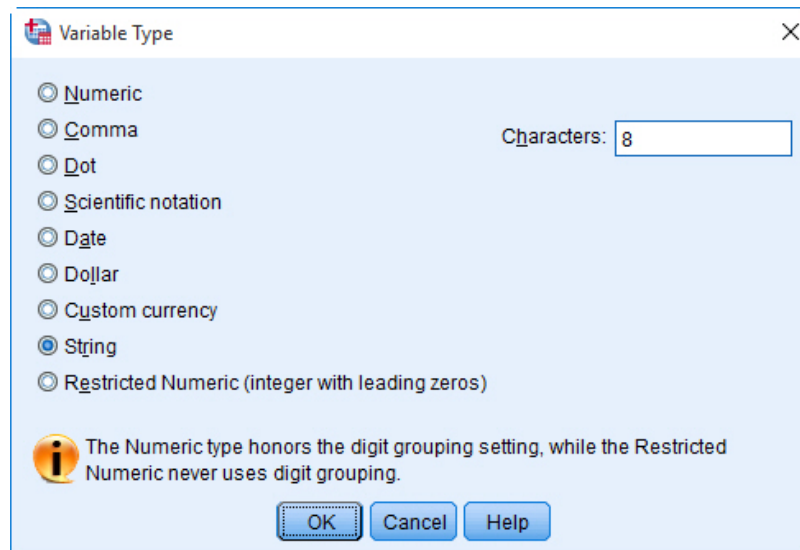


Figure 5 – Variable Type Dialog Box

5. Type **Sex** in row two under the **Name** column, and then press the **Enter** key.
6. Click the cell in row two under the **Decimals** column, and then change the entry to **0** using the spin box.
7. Type **What is your sex?** in row two under the **Label** column, and then press the **Enter** key.
8. Click **None** in row two under the **Values** column, and then click the **Ellipses** button.
9. In the **Value Labels** dialog box, type **1** in the **Value** box, type **Female** in the **Label** box, and then click the **Add** button (see Figure 6).
10. Repeat step 9 using a value of **2** and a label of **Male**.
11. Click the **OK** button.

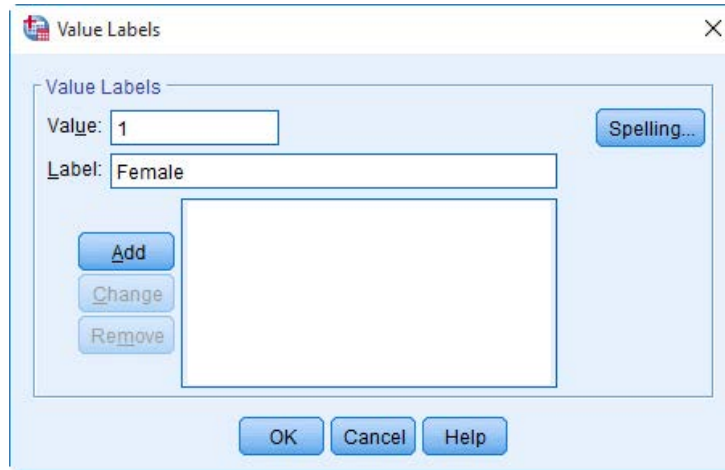


Figure 6 – Value Labels Dialog Box (Sex)

12. Type **GPA** in row three under the **Name** column, and then press the **Enter** key.
13. Type **Age** in row four under the **Name** column, and then press the **Enter** key.
14. Click the cell in row four under the **Decimals** column, and then change the entry to **0** using the spin box.
15. Type **What is your age?** in row four under the **Label** column, and then press the **Enter** key.
16. Click **None** in row four under the **Values** column, and then click the **Ellipses** button.
17. In the **Value Labels** dialog box, type **1** in the **Value** box, type **19 or Younger** in the **Label** box, and then click the **Add** button.
18. Repeat step 17 for values **2** through **5** and label them as shown in Table 3. See Figure 7 for the results.
19. Click the **OK** button.

Table 3 – Value Labels

Value	Label
1	19 or Younger
2	20-23
3	24-27
4	28-31
5	32 or Over

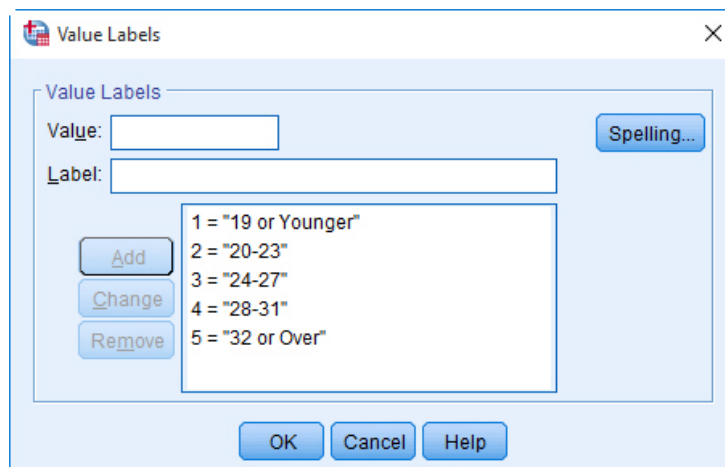


Figure 7 – Value Labels Dialog Box (Age)

Entering Data

After defining variables, you can enter data for each case. If variables are defined with the **Numeric** data type, then numeric data should be entered. SPSS Statistics only accepts numeric digits (0-9) for a Numeric data type. If variables are defined as **String** data, any keyboard character can be entered.

To enter data:

1. Click the **Data View** tab in the lower-left corner of the **Data Editor** window (see Figure 8).
2. Click in a cell and type the corresponding data. The entry is also displayed in the **Cell Editor** (see Figure 9).

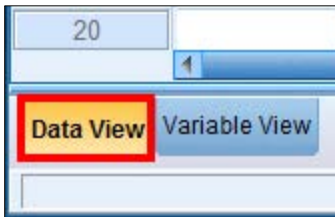


Figure 8 – Data View Tab

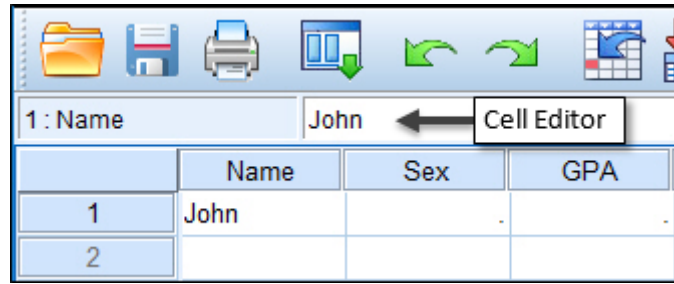


Figure 9 – Cell Editor

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).


Research Question # 1

What kind of device do people prefer to own?

Frequency Analysis

We can use frequency analysis to answer the first research question. **Frequency analysis** is a descriptive statistical method that shows the number of occurrences of each response chosen by the respondents. When using frequency analysis, SPSS Statistics can also calculate the mean, median, and mode to help users analyze the results and draw conclusions. The following example uses a frequency analysis to answer “Research Question # 1: What kind of device do people prefer to own?” using the data collected from the sample survey (see Appendix).

To perform frequency analysis:

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 **Part1.sav** file, and then click the **Open** button.
3. Click the **Analyze** menu, point to **Descriptive Statistics**, and then click **Frequencies** (see Figure 10).

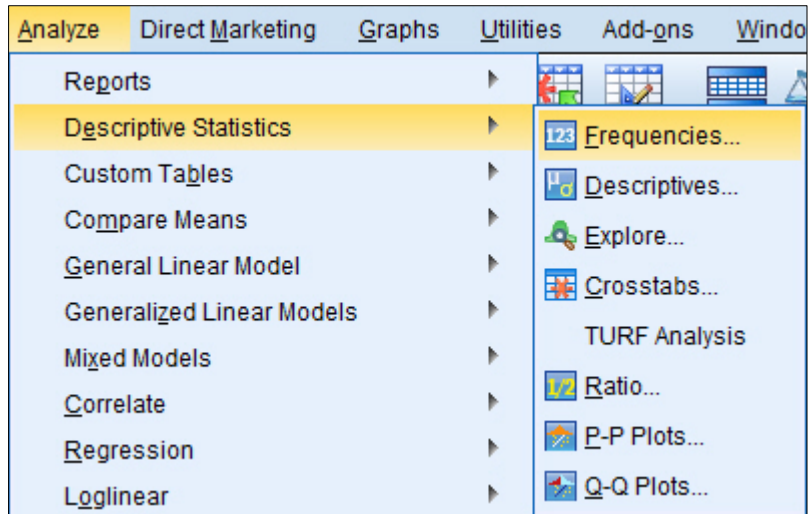



Figure 10 – Frequencies Selected on the Analyze Menu

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 (see Figure 11).
5. Make sure that the **Display frequency tables** check box is selected.
6. Click the **Statistics** button.

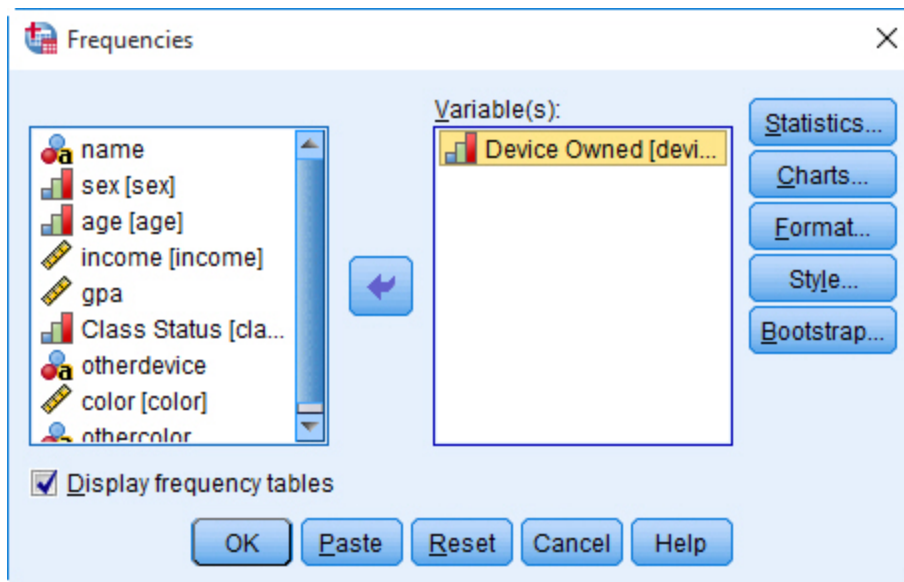


Figure 11 – Frequencies Dialog Box

7. In the **Frequencies: Statistics** dialog box, in the **Central Tendency** section, select the **Mean**, **Median**, and **Mode** check boxes (see Figure 12).
8. In the **Dispersion** section, select the **Std. deviation** check box.
9. Click the **Continue** button.

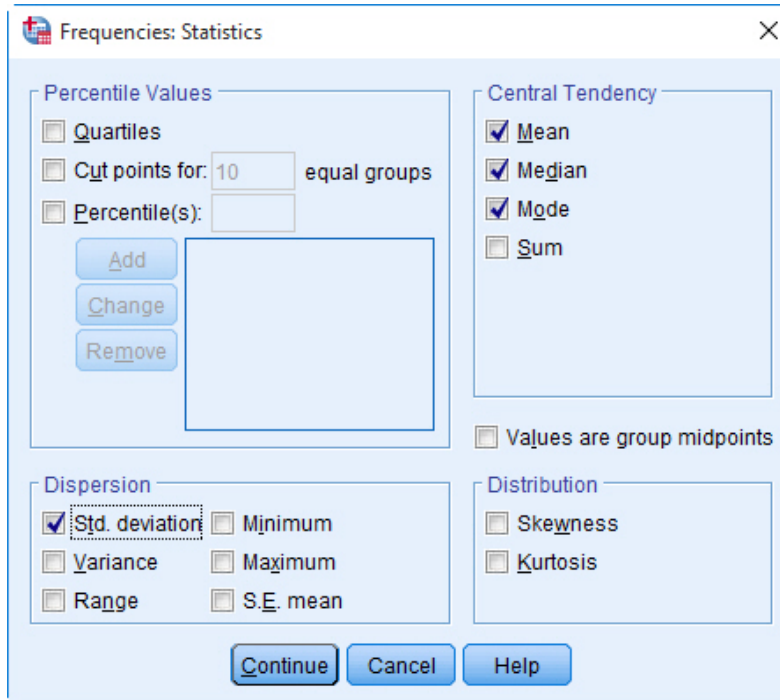


Figure 12 – Frequencies: Statistics Dialog Box

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 (see Figure 13).

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		

Figure 13 – Frequencies Output

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.

Answer to Research Question # 1 What kind of device do people prefer to own?

Answer: PC

Explanation: Look at question 7 in the Sample Survey. Notice that option 3 is PC. In the output **Statistics** table, the mode for **Device Owned** is 3, which is PC. In addition, the frequency analysis results for **Device Owned** indicate that 49 out of 80 people own a PC device. This can be considered their preference.


Research Question # 2 What color do people prefer for their device?

Crosstabs

Crosstabs are used to examine the relationship between two variables. To answer the second research question, users will analyze two variables: **Device Owned** and **Color** (which indicates color preference). Using crosstabs will show the intersection between these two variables and reveal the device type and color preferred by most people.

To perform a crosstabs analysis:

1. Switch to the **Data Editor** window.

NOTE: You can minimize the **Output Viewer** window by clicking the **Minimize** button  in the upper-right corner of the window.

2. In **Data View**, click the **Analyze** menu, point to **Descriptive Statistics**, and then click **Crosstabs** (see Figure 14).

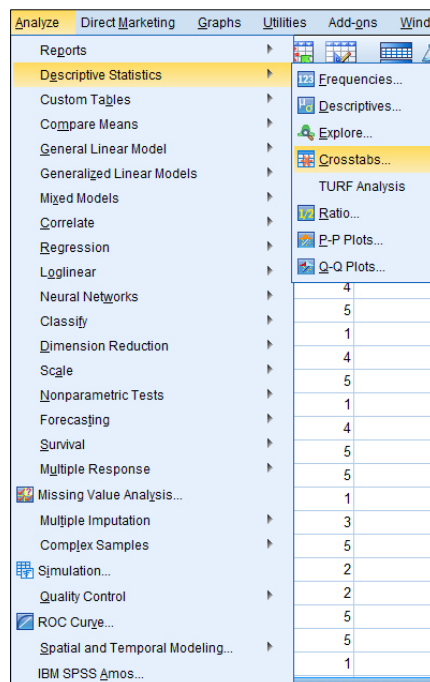


Figure 14 – Crosstabs Selected on the Analyze Menu

- In the **Crosstabs** dialog box, select the **Device Owned** variable in the left box, and then click the transfer arrow button to move it to the **Row(s)** box (see Figure 15).
- Select the **color** variable in the left box, and then click the transfer arrow button to move it to the **Column(s)** box.

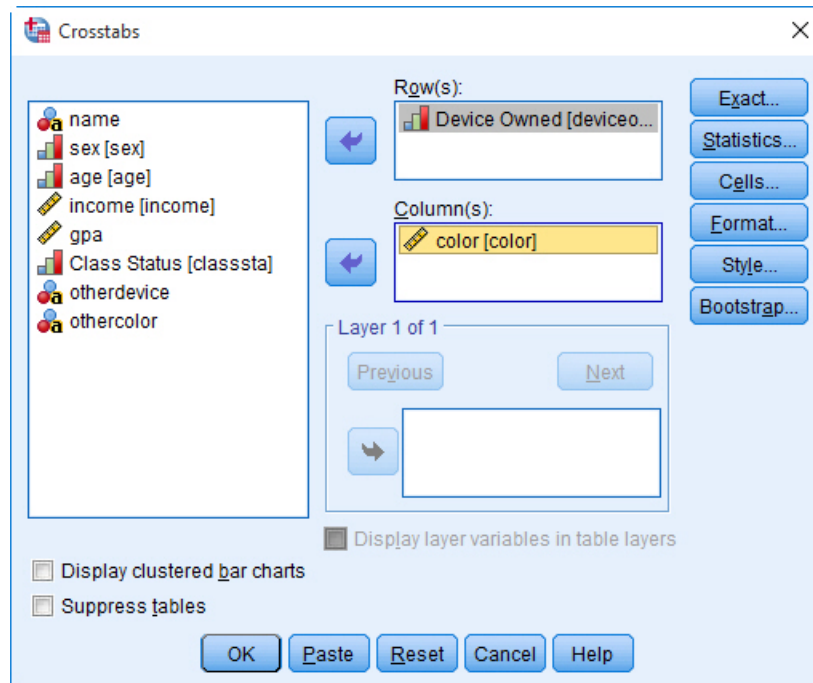


Figure 15 – Crosstabs Dialog Box

- Click the **OK** button. The **Output Viewer** window opens and displays two tables: **Case Processing Summary** and the **Crosstabulation** matrix (see Figure 16).

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Device Owned * color	73	91.3%	7	8.8%	80	100.0%

Device Owned * color Crosstabulation							
Count		color					Total
		black	silver	navy blue	white	5	
Device Owned	Tablet	2	0	1	0	0	3
	Mac	3	2	0	2	3	10
	PC	16	13	5	5	10	49
	Smartphone	3	0	0	0	0	3
	Other	2	2	1	2	1	8
Total		26	17	7	9	14	73

Figure 16 – Crosstabs Output

Answer to Research Question # 2
What color do people prefer for their device?

Answer: PC in black color

Explanation: As shown in the **Crosstabulation** matrix above, **PC** is the most preferred device type from the row variable (**Device Owned**). From the column variable (**color**), **black** is shown as the most preferred color. Therefore, you can conclude that most people prefer **PC** devices that are **black**.

Data Manipulation

Data files are not always organized to meet specific user needs. For example, users may wish to select a specific subject or split the data file into separate groups for analysis.

Select Cases

If you have two or more subject groups in your data and you want to analyze each subject independently, you can use the **Select Cases** option. For example, the data we are currently analyzing has both male and female participants. However, if you wish to analyze only female cases, then you select **sex** cases and set the condition for **female** cases only.

To select cases for analysis:

1. Switch to the **Data Editor** window.
2. In **Data View**, click the **Data** menu, and then click **Select Cases** (see Figure 17).

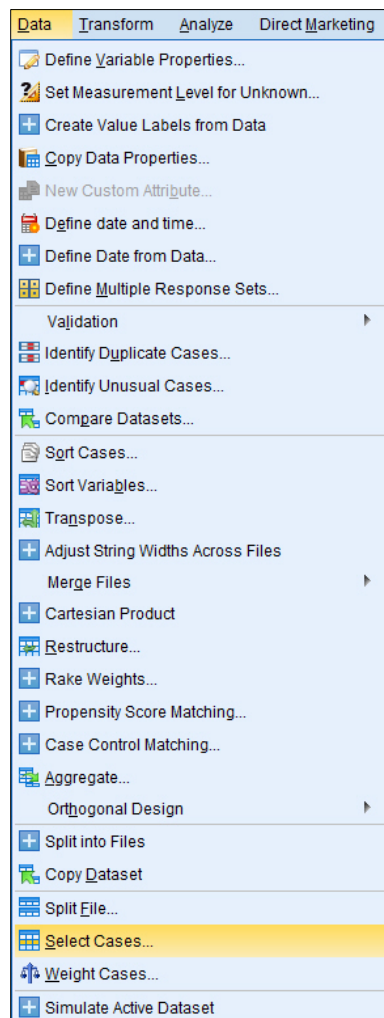


Figure 17 – Select Cases Selected on the Data Menu

- In the **Select Cases** dialog box, in the **Select** section, select the **If condition is satisfied** option button, and then click the **If** button (see Figure 18).

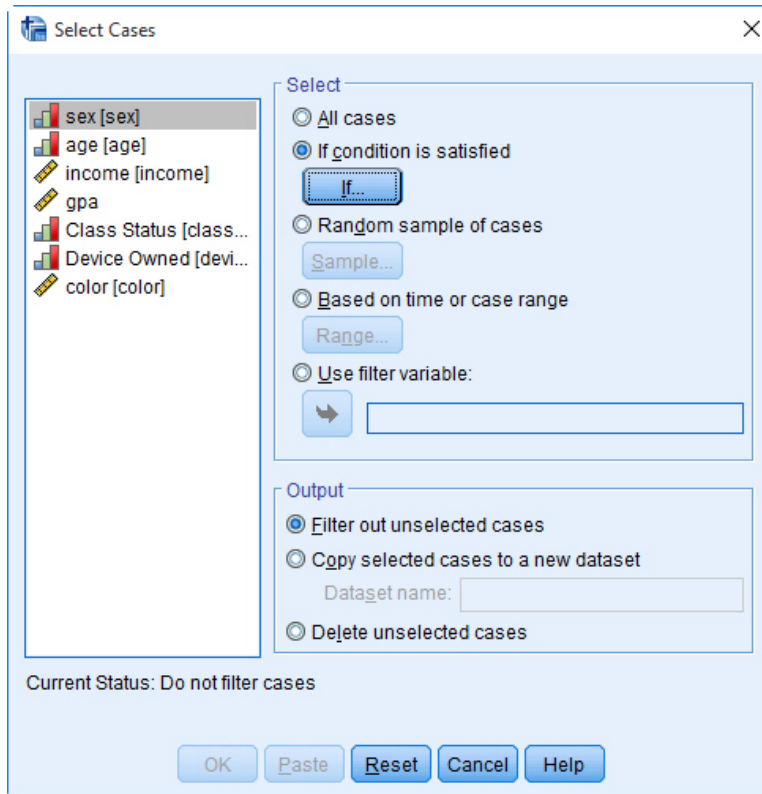


Figure 18 – Select Cases Dialog Box

- In the **Select Cases: If** dialog box, select the **sex** variable in the left box, and then click the transfer arrow button to move it to the right box (see Figure 19).
- Click the = button **=**, click the **1** button **1**, and then click the **Continue** button.

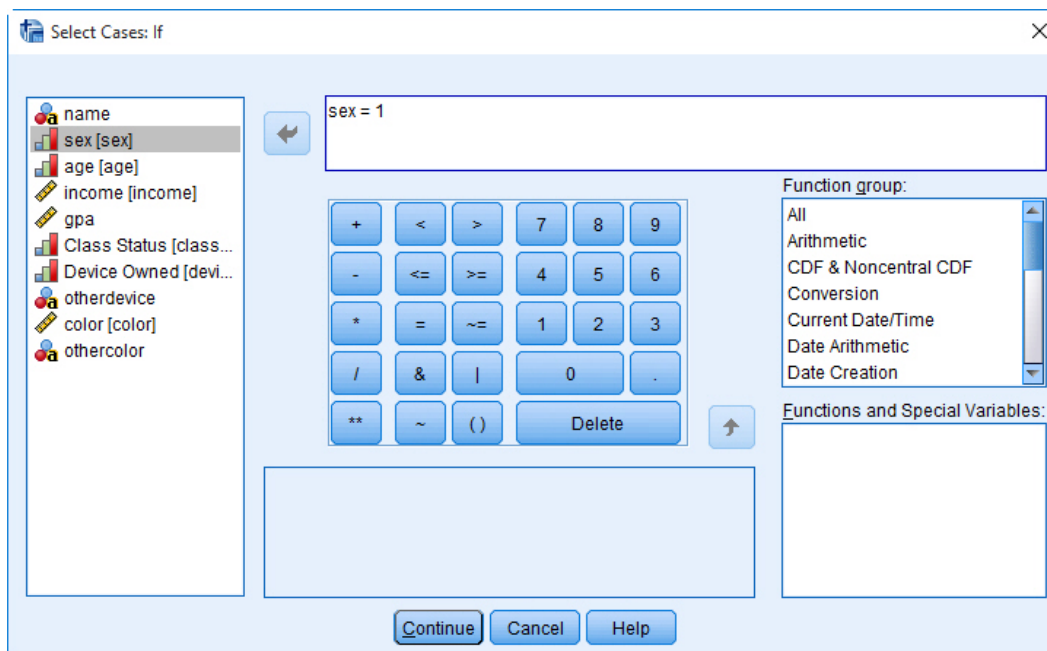


Figure 19 – Select Cases: If Dialog Box

6. In the **Select Cases** dialog box, click the **OK** button.
7. The **Output Viewer** window opens, indicating that all males are excluded from the statistical analysis. Minimize the **Output Viewer** window.

NOTE: In the **Data Editor** window, in **Data View**, the row numbers of the excluded cases are crossed out (see Figure 20).

	name	sex	age
1	Jones	2	5
2		2	2
3		2	3
4		2	4
5		2	3
6	Kidman	2	1
7	Jackson	1	4
8	Hopkins	1	4
9	Kooshian	1	1

Figure 20 – Data View Showing Excluded Cases

8. Rerun the crosstabs analysis by following steps 2 through 5 in the [Crosstabs](#) section of this handout. The **Output Viewer** window opens (see Figure 21).

In the Crosstabulation table of the Output Viewer window, look at the column for the most preferred color and the row for the device types. What is the device color most preferred by women? Ten women chose **PC** with color option **5**. Thus, you may conclude that most female participants prefer the color **5** for **PC** devices. However, what does **5** represent? This problem arose by not labeling the variable value **5** as **Other**. Moreover, the label **Other** does not indicate any particular color, making it difficult to draw a conclusion. In order to avoid such problems, it is suggested that you provide a blank space where participants can specify other color preferences besides the ones specified in the survey questionnaire.

Example:

What color do you like to have for your device?

1. Black 2. Silver 3. Navy Blue 4. White 5. Other _____

Case Processing Summary							
		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Device Owned * color		36	97.3%	1	2.7%	37	100.0%

Device Owned * color Crosstabulation						
		color				Total
		black	silver	white	5	
Device Owned	Mac	1	1	2	3	7
	PC	8	1	5	10	24
	Other	2	0	2	1	5
Total		11	2	9	14	36

Figure 21 – Select Cases Output

Research Question # 3

Is device color preference different between males and females?

Split File

To answer the third research question, we need to split the file. You can analyze one particular group of subjects using the **Select Cases** option. However, if you wish to compare the response or performance differences by groups within one variable, it is best to use the **Split File** option.

To split a file for analysis:

1. Switch to the **Data Editor** window.
2. In **Data View**, click the **Data** menu, and then click **Select Cases**.
3. In the **Select Cases** dialog box, in the **Select** section, select the **All cases** option button, and then click the **OK** button. This removes the select cases option set in the previous procedure.
4. The **Output Viewer** window opens, indicating that all cases are included. Minimize the **Output Viewer** window.
5. Click the **Data** menu, and then click **Split File** (see Figure 22).
6. In the **Split File** dialog box, select the **sex** variable in the left box (see Figure 23).
7. Select the **Compare groups** option button.
8. Click the transfer arrow button to move the **sex** variable to the **Groups Based on** box.
9. Click the **OK** button.

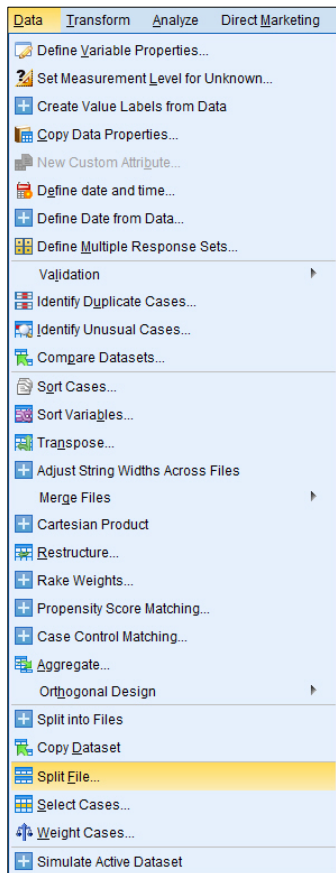


Figure 22 – Split File Selected on the Data Menu

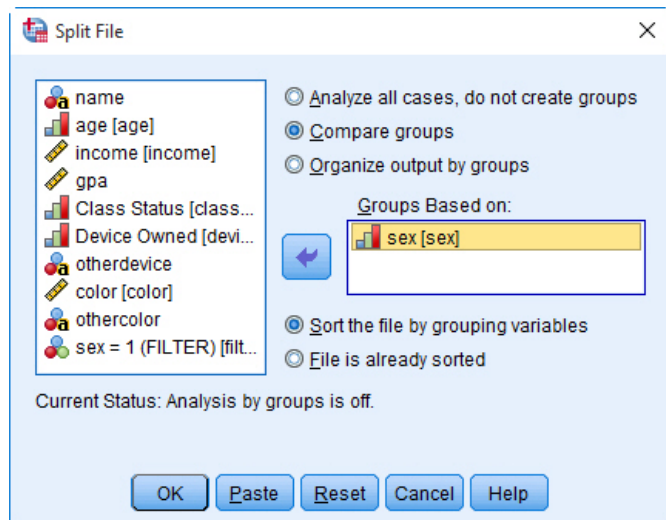


Figure 23 – Split File Dialog Box

10. The **Output Viewer** window opens, indicating that the file is split by sex. Minimize the **Output Viewer** window.
11. Rerun the crosstabs analysis by following steps 2 through 5 in the [Crosstabs](#) section of this handout. The **Output Viewer** window opens and displays the Crosstabulation table (see Figure 24).

Count			color					Total
sex			black	silver	navy blue	white	5	
Female	Device Owned	Mac	1	1		2	3	7
		PC	8	1		5	10	24
		Other	2	0		2	1	5
	Total		11	2		9	14	36
Male	Device Owned	Tablet	2	0	1			3
		Mac	2	1	0			3
		PC	8	12	5			25
		Smartphone	3	0	0			3
		Other	0	2	1			3
	Total		15	15	7			37

Figure 24 – Split File Output Data

Answer to Research Question # 3

Is device color preference different between males and females?

Answer: Yes

Explanation: There is a device color preference difference based on sex. From the Crosstabulation output, females prefer **PC of other** color over the colors black, silver, navy blue, and white. The male group prefers **silver PC** devices.

Find and Replace

The **Find and Replace** function is useful for tasks such as updating a respondent's name. Users can use **Find and Replace** in Data View. However, in Variable View, only the **Find** function is available.

To use the Find and Replace function:

1. Switch to the **Data Editor** window.
2. In **Data View**, click to select any cell in the column to be searched or click the column header. In this case, click a cell in the **Name** column.
3. Click the **Edit** menu, and then click **Replace**. The **Find and Replace** dialog box opens with the **Replace** tab selected (see Figure 25).
4. In the **Find** box, type **Clinton**.
5. In the **Replace with** box, type **Cliff**.
6. Click the **Show Options** button.
7. Under **Match to**, select the **Entire cell** option button.

NOTE: Under the **Match to** section of the **Find and Replace** dialog box, **Contains** means SPSS Statistics will find each instance of the word/phrase/number appearing in a cell, whether or not it is the only information enclosed. The **Entire cell** option will find the word/phrase/number that matches the entire cell as a whole. Selecting the **Begins with** or **Ends with** option will search the character indicated by the user.

8. Click the **Replace All** button.

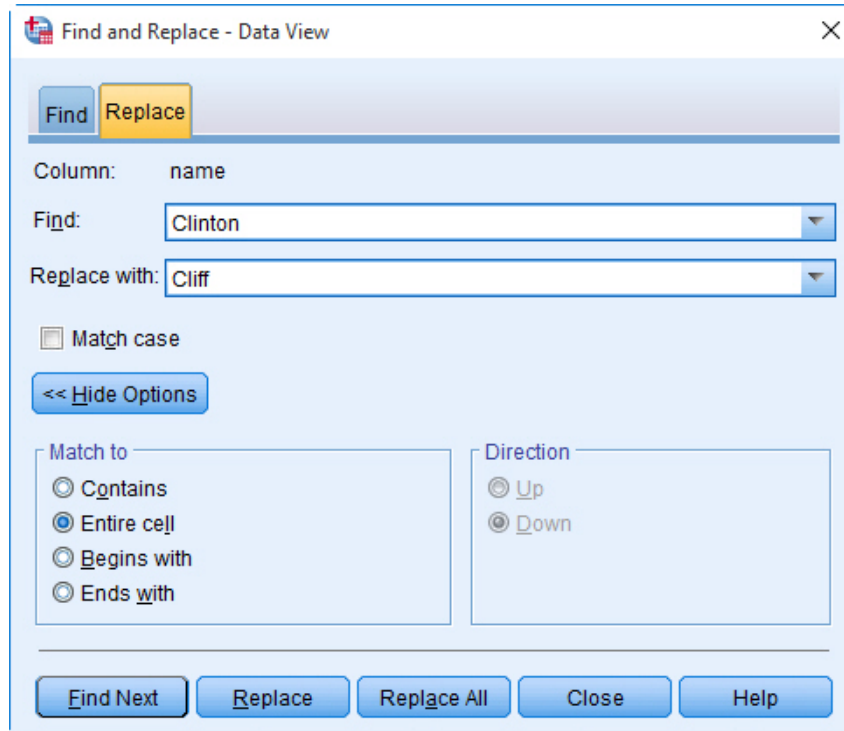


Figure 25 – Find and Replace Dialog Box in Data View

9. A message dialog box opens, indicating the number of replacements made. Click the **OK** button (see Figure 26).

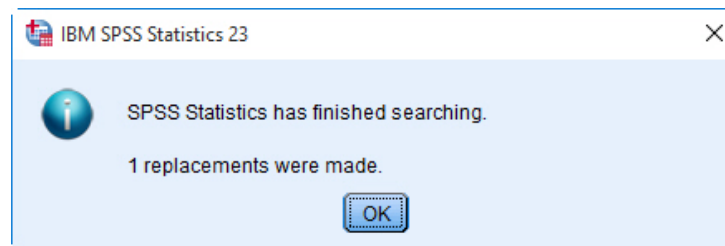


Figure 26 – Message Dialog Box

10. In the **Find and Replace** dialog box, click the **Close** button.

Reports

When the statistical analysis is complete, the final step is to create a report that may include SPSS Statistics output (e.g., graphs and tables) for supporting your analysis. Using the **Copy** and **Paste** functions, the tables/graphs generated in SPSS Statistics can be copied from the **Output Viewer** window and pasted into a **Microsoft Word** document without having to create new tables or graphs.

To create a report using Microsoft Word:

1. In the **Output Viewer** window, right-click a table. A box appears around the table and a red arrow → appears to the left of the table to indicate that it is selected.
2. Click **Copy** on the shortcut menu.
3. Start **Microsoft Word** and create a new document.
4. Right-click in the Word document, and then click **Paste** on the shortcut menu. The table is copied into the Word document.

Appendix

SAMPLE SURVEY

Research Questions

1. What kind of device do people prefer to own?
2. What color do people prefer for their device?
3. Is device color preference different between males and females?

Survey Questions

1. What is your name? _____
2. What is your sex? _____
3. What is your GPA? _____
4. What is your age?
 1. 19 or Younger
 2. 20-23
 3. 24-27
 4. 28-31
 5. 32 or Over
5. How much do you make in a month?
 1. Less than \$1000
 2. \$1000-\$1499
 3. \$1500-\$1999
 4. \$2000-\$2499
 5. Over \$2500
6. What is your class standing?
 1. Freshman
 2. Sophomore
 3. Junior
 4. Senior
 5. Graduate
7. What kind of device do you own?
 1. Tablet
 2. Mac
 3. PC
 4. Smartphone
 5. Other: _____
8. What color do you like to have for your device?
 1. Black
 2. Silver
 3. Navy Blue
 4. White
 5. Other: _____