

What are some best practices for working with “Check All That Apply” survey data (multiple response sets)?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are some best practices for working with “Check All That Apply” survey data (multiple response sets)?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149703>

"Check All That Apply" survey data, also known as multiple response sets, require specific best practices to ensure accurate and meaningful analysis. Firstly, it is crucial to clearly define the purpose and scope of the survey to avoid confusion for respondents when selecting multiple options. Additionally, having a balanced and exhaustive list of response options can prevent bias and aid in data organization. Furthermore, using a consistent and logical ordering of options can make the data more interpretable. Other best practices include providing a "prefer not to answer" option and properly formatting the data for analysis. Lastly, validating the survey results through cross-tabulations and statistical tests can ensure the reliability and validity of the data. Overall, following these best practices can lead to more accurate and insightful analysis of "Check All That Apply" survey data.

Introduction

This tutorial is a primer on how to work with data from multiple choice, multiple-response (or "check all that apply") questions in SPSS Statistics.

Jump to:

[Introduction: About Multiple Response Set Variables](#)[Data Set-Up](#)[Sample Data for This Tutorial](#)[Counting the Number of Selected Options using Count Values Within Cases](#)[Defining Multiple Response Sets in SPSS](#)[Example: Multiple Response Frequency Tables](#)[Example: Multiple Response Crosstabs](#)

Multiple response sets occur when you have a set of related choices or characteristics in which a subject or experimental unit can possess one or more of those characteristics. In this tutorial, we will focus on a specific type of multiple response set: multiple response (or "check-all-that-apply") questionnaire items.

A multiple response question presents a list of possible answer options, and the respondent selects all options that are true for them. For example, suppose we are interested in surveying a group about what types of electronic devices they own, and suppose we are especially interested in the three most common types of mobile computing devices: laptops, phones, and tablets. We might create a survey question like this one:

Which of the following electronic devices do you own?

Check all that apply.

- Laptop
- Phone
- Tablet
- Other

As individual users complete the survey, their selections might look like this:

User 1

Selects "laptop" and "phone" and "tablet"

- Laptop
- Phone
- Tablet
- Other

User 2

Selects "tablet"

- Laptop
- Phone
- Tablet
- Other

User 3

Selects "phone" and "other"; types "mp3 player" in the write-in box

- Laptop
 Phone
 Tablet
 Other

This particular question type is deceptively simple. On its surface, it looks similar to "single-choice" multiple choice questions, which can be summarized using (univariate) frequency tables. However, this is not the case for multiple response questions: each checkbox functions like a "Yes or No" question. For example, we could restructure this question into a series of single-choice, "Yes or No" questions:

Do you own any of the following electronic devices?

	Yes	No
Laptop	<input type="radio"/>	<input type="radio"/>
Phone	<input type="radio"/>	<input type="radio"/>
Tablet	<input type="radio"/>	<input type="radio"/>
Other <input type="text"/>	<input type="radio"/>	<input type="radio"/>

This means that **one multiple-response question is actually composed of several binary variables**. (There will be as many binary variables as there are "selectable" options.) Additionally, we are not just concerned about how many individuals selected a given choice; we may also care about how many of the options were selected, and what combinations of the options were most common (i.e., are the selections correlated).

To properly analyze these responses, our data must be structured correctly. In practice, there are two basic data structures for this type of data, but one of them is much easier to work with than the other.

Data for this question is recorded in a **single column**. The person's selections are written as text,

with commas (or other delimiter characters) between their choices. Data for this question is recorded in **multiple columns**, with one column per answer option. If a person selected that option, they are assigned a "1" for that variable; if they did not select that option, their data value is left blank (or is assigned a number code indicating non-selection).

Single-column structure

	B
1	What of the following electronic devices do you own? Check all that apply.
2	Laptop, Phone, Tablet
3	Phone, Tablet
4	mp3 player
5	Phone, Tablet
6	Phone
7	Laptop, Phone
8	don't own any of these
9	Laptop, Tablet
10	Laptop, Phone, Tablet, digital camera

Multiple-column structure

	C	D	E	F	G
1	owns_laptop	owns_phone	owns_tablet	owns_other	owns_other_TEXT
2	1	1	1		
3		1	1		
4				1	mp3 player
5		1	1		
6		1			
7	1	1			
8				1	don't own any of these
9	1		1		
10	1	1	1	1	digital camera

If you are given the choice between these two structures, **the multiple-column scheme is strongly preferred**. If your data is recorded using the single-column structure, you will need to "clean up" the data to get it into the one-column-per-selection format.

Data Set-Up

To properly analyze multiple response questions in SPSS, your dataset should have the following structure:

Each row (case) should represent one subject, survey response, or experimental unit. For a given multiple response question, each answer option should be represented in a separate column (variable).

The multiple response variables should be numeric. If they are string, you will need to convert them to numeric codes (see the [Automatic Recode](#) procedure). The data values should follow one of these two schemes:






Numeric code (typically 1) if present, blank (missing) if not present. Numeric codes representing present and not present (such as 0=Absent, 1=Present).

The following two examples demonstrate both schemes using the same underlying data. In these examples, the columns represent the answers to a check-all-that-apply question, "Which of the following devices do you own?", with four answer options: laptop, phone, tablet, or "other". In plain language, the data used in both examples:

Cases 1, 2, and 5 own a laptop *and* a phone *and* a tablet. Cases 3, 4, 6, and 8 own a laptop *and* a phone. Case 7 owns only a phone.

Numeric code if present, blank if absent






In this coding scheme, we have a distinct numeric code representing the "checked" or "present" state, but use a missing value (blank) to represent the "unchecked" or "absent" state.

	 ids	 owns_laptop	 owns_phone	 owns_tablet	 owns_other
1	27478	1	1	1	.
2	33802	1	1	1	.
3	43044	1	1	.	.
4	42396	1	1	.	.
5	26214	1	1	1	.
6	38034	1	1	.	.
7	41636	.	1	.	.
8	22773	1	1	.	.

In this example, 1 denotes "present" or "checked", and blank cells denote "absent" or "not checked". (Remember that when entering data, we do not type anything in the cell when the value is missing. The dot you see in the cell is something that SPSS displays, not something that you as the user add.) Variable labels are strongly recommended, since those determine the labeling used in the multiple response frequency tables. Value labels are also useful, but are not a strict requirement.

Numeric codes to indicate present/not present

In this coding scheme, we have a distinct numeric code representing the "checked" or "present" state, and a distinct numeric code representing the "unchecked" or "absent" state.

	 ids	 owns_laptop	 owns_phone	 owns_tablet	 owns_other
1	27478	1	1	1	0
2	33802	1	1	1	0
3	43044	1	1	0	0
4	42396	1	1	0	0
5	26214	1	1	1	0
6	38034	1	1	0	0
7	41636	0	1	0	0
8	22773	1	1	0	0

In this example, 0 denotes "absent" or "not checked", and 1 denotes "present" or "checked". You do not necessarily have to use the numbers 0 and 1, but you should use the same numeric codes across all of the columns. Value labels are strongly recommended, so that you can remember the

meanings of the codes.

Sample Data for This Tutorial

In this tutorial, we will be using simulated data from a hypothetical survey with two questions. Neither question was required, so respondents could choose to skip one or both questions.

Which of the following electronic devices do you own?

Check all that apply.

- Laptop
- Phone
- Tablet
- Other

What is your gender?

- Male
- Female

The data for the multiple response question was encoded using the 1s/blanks scheme, and the answers to the question on gender were encoded as 0=Male, 1=Female.

[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.csv](#)

[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.sav](#)

Counting Selected Choices and Identifying Non-respondents

An important task when working with check-all-that-apply questions is being able to say how many people did *not* answer the question. This task is not as straightforward as it is with single-choice multiple-choice questions, where we can simply count the number of missing values in a single column. That approach will not work with multiple-response questions, because the answers are spread across multiple variables, and can be selected independently. For example, someone who responds that they own a phone will still have missing values for laptop and tablet and other. This person clearly answered the question, despite having "missing values" on some of the variables in the set. It stands to reason that a person who did not answer the question must have missing

values on **all** variables in the response set. How do we count the number of nonmissing responses a person gave?

Additionally, we may want to know how many options respondents tended to select. Did most people only select 1 option, or did most people tend to select 2 or 3 options?

We can answer both of these questions using the **Count Values Within Cases** procedure in SPSS. This procedure takes a set of variables and counts the number of times a specific value occurs for a given case/row. This "count" is added as a new variable to the dataset, which we can then use to apply filters.

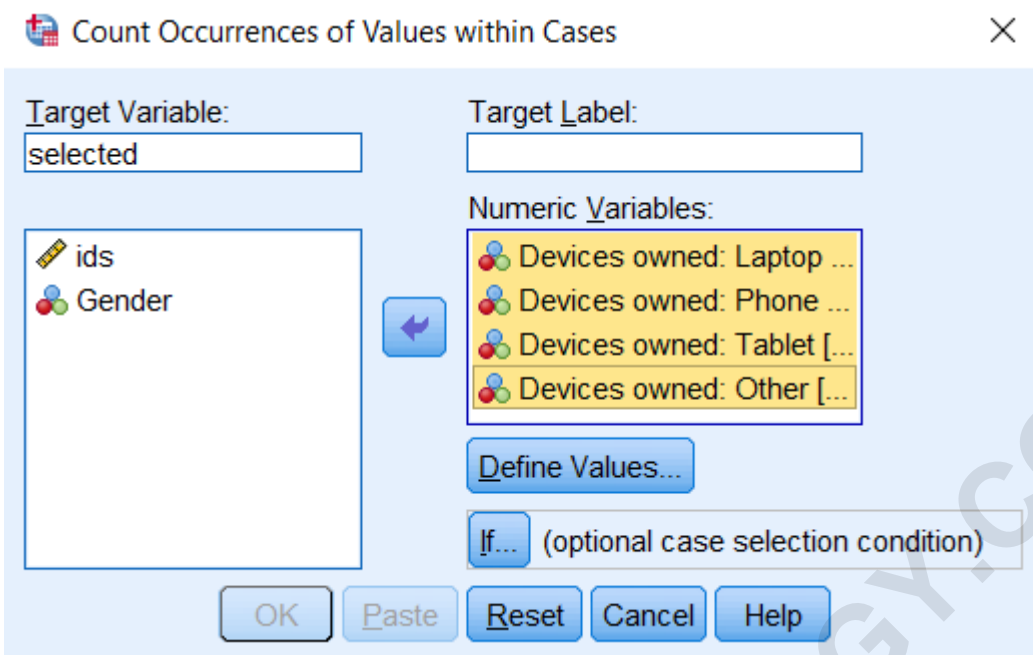
In our example data, we used the number 1 to indicate "present", so we want to count the number of 1's a person has across the four multiple response variables. We can use Count Values Within Cases to count the number of "checked boxes" for a given respondent. If someone does not have any 1's, they will have a count of 0.

Count Values Within Cases can be configured to count any number or range of numbers, and can even count missing values. In this example, we choose to count the number of 1's, so individuals who selected zero choices will have values of 0, and individuals who answered the question will have counts greater than 0. This interpretation is more intuitive, and makes it easy to filter out non-responders.

Running the Procedure

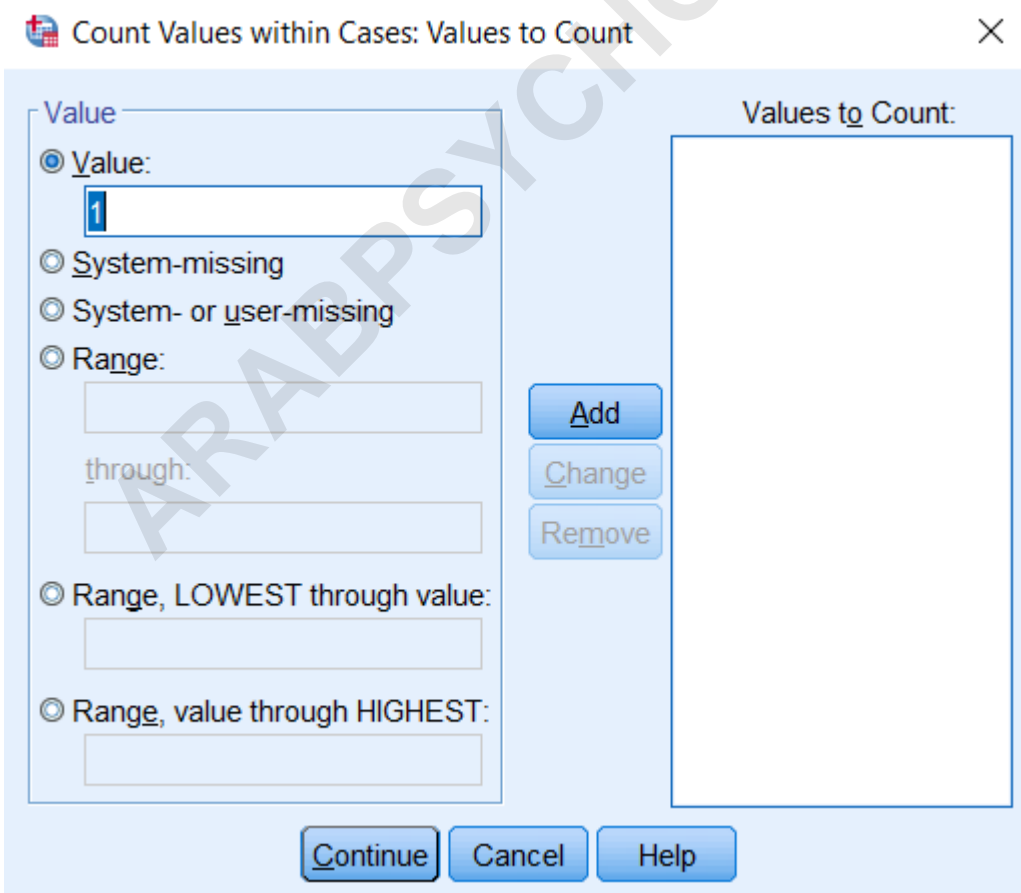
Using the Dialog Windows

Click **Transform > Count Values within Cases**. In the Target Variable box, type a name for the new variable to be created. Let's call our new variable selected. Double click on the variables owns_laptop, owns_phone, owns_tablet, and owns_other in the left column to move them to the Variables box.



Click **Define**

Values. In the left column, type the number 1 in the **Value** box, then click **Add**. You should see the number 1 added to the Values to Count column.



Click **Continue** to save the change. Click **OK**.

Using Syntax

```
COUNT selected=owns_laptop owns_phone owns_tablet owns_other(1).  
EXECUTE .
```

In this syntax:

`COUNT` is the name of the procedure. The name of the new variable to be created, `selected`, appears to the left of the equals sign. After the equals sign, we list the names of all variables to count. We use spaces between the variable names. After the name of the last variable, we put the value to count in parentheses. In this case, we are counting the value 1. The statement ends with a period. The `EXECUTE` statement tells SPSS to carry out the computation and write the result to the active dataset. (If you run the `COUNT` statement without the `EXECUTE` statement, SPSS will "queue up" the command, but not actually carry it out.)

Output

The Output window will display the syntax from the Count Values within Cases command, but will not show any table output. To see the result, go into the Data Editor window; if we were successful, our new variable should appear at the end of the dataset (you may need to scroll to the right to see it).

The screenshot shows the IBM SPSS Statistics interface. The title bar reads '*Sample Dataset 2019 - Multiple Response Question with Categorical Variable.sav [D...'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data view with 8 cases and 6 variables. The variables are: ids, owns_laptop, owns_phone, owns_tablet, owns_other, and selected. The 'selected' variable is highlighted with a red box. The data is as follows:

	ids	owns_laptop	owns_phone	owns_tablet	owns_other	selected
1	27478	1	1	1	.	3.00
2	33802	1	1	1	.	3.00
3	43044	1	1	.	.	2.00
4	42396	1	1	.	.	2.00
5	26214	1	1	1	.	3.00
6	38034	1	1	.	.	2.00
7	41636	.	1	.	.	1.00
8	22773	1	1	.	.	2.00

The interface also shows 'Data View' and 'Variable View' tabs, and a status bar at the bottom indicating 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'.

Notice how cases 1, 2, and 5 had values of 1 for owns_laptop, owns_phone, and owns_tablet, and that their value of selected is 3. Cases 3, 4, 6, and 8 had values of 1 for owns_laptop and owns_phone, so their value of selected is 2. Case 7 only had a 1 for owns_phone, so their value of selected is 1.

We can now look at how many devices the respondents owned by creating a frequency table of variable selected (**Analyze > Descriptive Statistics > Frequencies**).

Statistics

selected

N	Valid	435
	Missing	0

selected

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	6	1.4	1.4	1.4
	1.00	34	7.8	7.8	9.2
	2.00	240	55.2	55.2	64.4
	3.00	143	32.9	32.9	97.2
	4.00	12	2.8	2.8	100.0
	Total	435	100.0	100.0	

From this table, we can see that six (6) respondents did not select any electronic devices. Keep this number in mind when reviewing the Multiple Response Frequencies output in the next example.

This table also tells us:

Thirty-four (34) respondents, or 7.8% of the sample, own a single electronic device. Two-hundred forty (240) respondents, or 55.2% of the sample, own two electronic devices. One-hundred forty-three (143) respondents, or 32.9% of the sample, own three electronic devices. Twelve (12) respondents, or 2.8% of the sample, own four electronic devices (i.e., selected all four answer options).

To filter out individuals who did not answer the multiple response question, use the Select Cases procedure to keep cases if `selected > 0` (selected greater than 0).

Defining Multiple Response Sets in SPSS

SPSS has a two-step process to use multiple response sets using the dialog windows:

Define the multiple response set.

Identify the variables representing the values for that set. Indicate which number code(s) should be counted as "present". Run the multiple response frequencies or crosstabs procedures.

After a multiple response set is defined, it is only retained as long as the SPSS session is active. Once you close SPSS, the multiple response set definition is erased; the next time you start SPSS,

A Variables in Set: The variables from the dataset that compose the multiple response set. For surveys, this is typically the set of columns corresponding to the "selectable" choices for a single survey question.

B Variables Are Coded As: The data values used to indicate that the category was present.

Dichotomies: Use if a single numeric value was used across all of the variables to indicate if the category was "present". **Categories:** Use if there was a range of number codes used to indicate if the category was present, or if there is more than one category that will be counted as "present".

Note that it is only possible to choose one of these schemes. If your data does not match one of these schemes, you may need compute recoded versions of the variables using the Recode into Different Variable procedure.

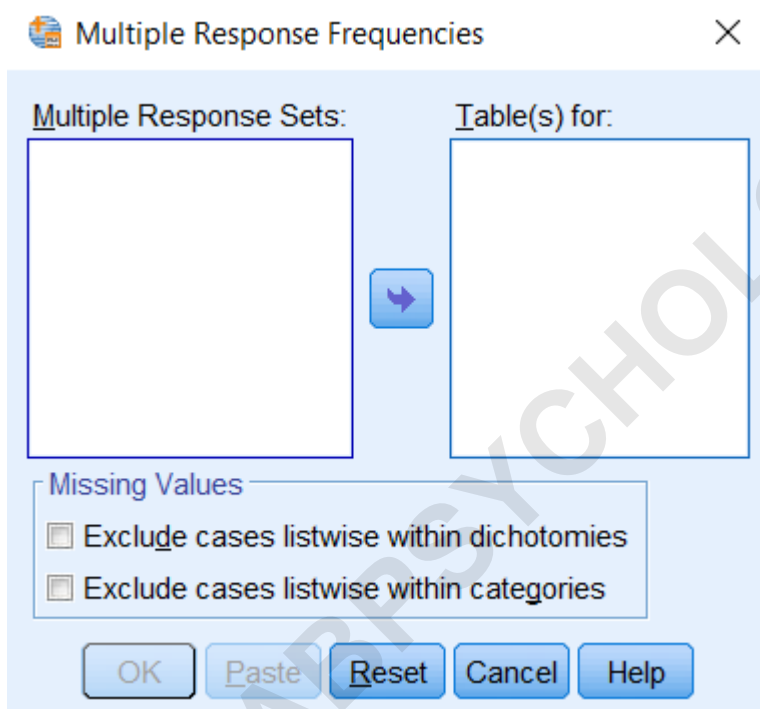
C Name and Label: The name (required) and label (optional) of the multiple response set. The

naming rules for multiple response set names are the same as the normal variable naming rules in SPSS (no spaces, must start with a letter).

DMultiple Response Sets: List of all response sets that have been defined in the current SPSS session. This panel will be blank if no response sets are defined.

Step 2: Multiple Response Frequencies

After setting up a multiple response set, you will be able to access the Multiple Response Frequencies option through the menus. To do this, click **Analyze > Multiple Response > Frequencies**.



All multiple response sets you've defined during the current SPSS session will appear on the left.

The two options in the Missing Values section control how cases with missing values should be treated. These settings will have different effects, depending on whether you use blanks versus numeric codes to represent unselected choices, and whether you specified a dichotomy or a range of category codes in the previous step:

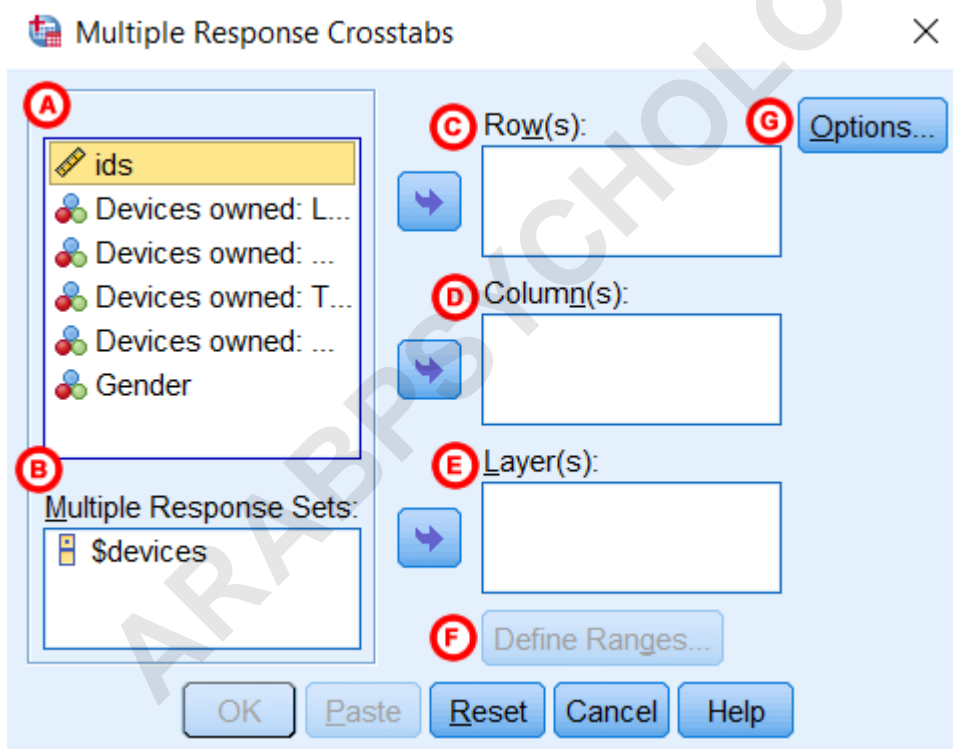
The Exclude cases listwise within dichotomies option will treat cases with any missing values as fully missing. If you coded your selected values as 1 and blank if not selected, this particular option will only count cases where all values were present. This particular option should only be used if you coded selected values as 1 and unselected values as 0 (or some other nonmissing numeric

code). The Exclude cases listwise within categories option will only consider a case as "missing" if it does not have at least one variable with the specified number code.

To avoid having to re-define the same response set, we recommend using the Paste button (instead of the OK button) to generate the command syntax code for the multiple response frequency table or crosstab. This is because the syntax command for multiple response sets, `MULTI RESPONSE`, contains the definition of the set in the command. (This will be illustrated in the example below.) Using the Paste button will write the syntax commands to the syntax window, which you can then use to execute the analysis without needing to go through the dialog windows.

Step 3: Multiple Response Crosstabs

After setting up a multiple response set, you will be able to access the Multiple Response Crosstabs option through the menus. To do this, click **Analyze > Multiple Response > Crosstabs**.



A Variable list: The variables in the current dataset. Categorical variables in this list can be used as Row, Column, or Layer variables. For each variable in this list that you use in the table, you will need to use the Define Ranges button to tell SPSS which number categories you want to be included in the table.

B Multiple Response Sets: The multiple response sets that have been defined during the current

session. These variables can be used as Row, Column, or Layer variables.

CRows: The variable(s) you want to be used as the rows in the crosstab.

DColumns: The variable(s) you want to be used as the columns in the crosstab.

ELayers: The variable(s) you want to be used as the "layer" variable in the crosstab. The categories of the layer variable will appear on the outermost edge of the table.

		Col 1	Col 2
Layer 1	Row 1		
	Row 2		
Layer 2	Row 1		
	Row 2		

If multiple variables are entered in the Row, Column, and/or Layer boxes, there will be a separate table for each unique combination of the row*column*layer variables.

FDefine Range: Opens the Define Range prompt. This option becomes available when you've added a regular variable to the Row, Column, or Layer box, and have clicked on the variable so that it's highlighted.

Multiple Response Cros... X

Minimum:

Maximum:

Continue Cancel Help

Note that this means that you cannot use string variables in these tables, and the numeric category codes you want to include in the table must be sequential (i.e., if you had categories 1=Disagree, 2=Neutral, 3=Agree, and you only wanted to include categories 1 and 3 in the table, you would need to recode the variable so that Neutral is not within the range.) If there are numbers between the minimum and maximum that are not represented in the observed data, those numbers will be ignored.

GOptions: Opens the Options window:

Multiple Response Crosstabs: Options

Cell Percentages

Row Column Total

Match variables across response sets

Percentages Based on

Cases Responses

Missing Values

Exclude cases listwise within dichotomies

Exclude cases listwise within categories

Continue Cancel Help

Cell Percentages: Add percentages to the table cells (in addition to showing the counts).

Row: Percentages will be based on the row total. **Column:** Percentages will be based on the column total. **Total:** Percentages will be based on the table total. **Match variables across response sets:** Applies only when the multiple response set definition used category code ranges. "Pairs the first variable in the first group with the first variable in the second group, and so on. If you select this option, the procedure bases cell percentages on responses rather than respondents. Pairing is not available for multiple dichotomy sets or elementary variables." **Percentages Based on:** If one or more of the "cell percentages" options are selected, controls the values used in the denominators of those calculations.

Cases: The marginal totals represent the number of cases in that group. When summing over the categories of a multiple response set, the sum of the cells may not equal the marginal total. **Responses:** The marginal totals equal the sum of the cells in the table. **Missing Values:** Change how missing values are handled in the table. By default, a case must have missing values on all response set variables to be counted as missing. Applying one of these settings will instead use listwise missing data handling; i.e., if a case has at least one missing value among the response set variables, it will be treated as missing. If your data uses the numeric/blank encoding scheme, applying these settings will generally lead to empty or mostly-empty tables.

Exclude cases listwise within dichotomies: Applies only when the multiple response set definition used dichotomies. **Exclude cases listwise within categories:** Applies only when the multiple response set definition used category code ranges.

How do I save multiple response sets defined through the menu system? - IBM

IBM SPSS Statistics Knowledge Base.
https://www.ibm.com/support/knowledgecenter/en/SSLVMB_26.0.0/statistics_mainhelp_ddita/spss/base/idh_mulc_opt.html

Example: Creating a Multiple Response Frequency Table

Problem Statement

Suppose we want to know what types of electronic devices (laptops, smartphones, and tablets) college students commonly own. Our desired summary would look something like this:

Devices owned	n	% of respondents (n=??)
Laptop		
Phone		
Tablet		
Other		

If we were to try to use the regular Frequencies procedure on this data (**Analyze > Descriptives > Frequencies**), the resulting tables would not be succinct:

Statistics

		Devices owned: Laptop	Devices owned: Phone	Devices owned: Tablet	Devices owned: Other
N	Valid	397	386	168	40
	Missing	38	49	267	395

Frequency Table

Devices owned: Laptop

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	397	91.3	100.0	100.0
Missing	System	38	8.7		
Total		435	100.0		

Devices owned: Phone

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	386	88.7	100.0	100.0
Missing	System	49	11.3		
Total		435	100.0		

Devices owned: Tablet

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	168	38.6	100.0	100.0
Missing	System	267	61.4		
Total		435	100.0		

Devices owned: Other

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	40	9.2	100.0	100.0
Missing	System	395	90.8		
Total		435	100.0		

The first table shows the number of valid and missing responses for each variable. Notice the

number of missing responses for each variable: Because we are using the scheme of 1=checked, missing=not checked, the missing values here actually represent the number of people who did not select that option. It does not necessarily mean that they did not answer the question! We should only consider individuals who left all four options blank as skipping the question. It's not possible to determine how many individuals left all four options blank from the basic Frequencies procedure.

In the individual frequency tables, we see the number of people who checked that option (in the rows labeled "Valid - 1"). The Percent column represents the proportion of the total sample who checked that option. Because this procedure can't determine if there were individuals who did not answer the question, we don't know for certain if we should use the total sample size as the denominator to compute the percentages.

Instead, we should use the Multiple Response Frequencies procedure, which can deal with all of these issues, and produce a table structured like the above.

Sample Dataset for This Example

[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.csv](#)

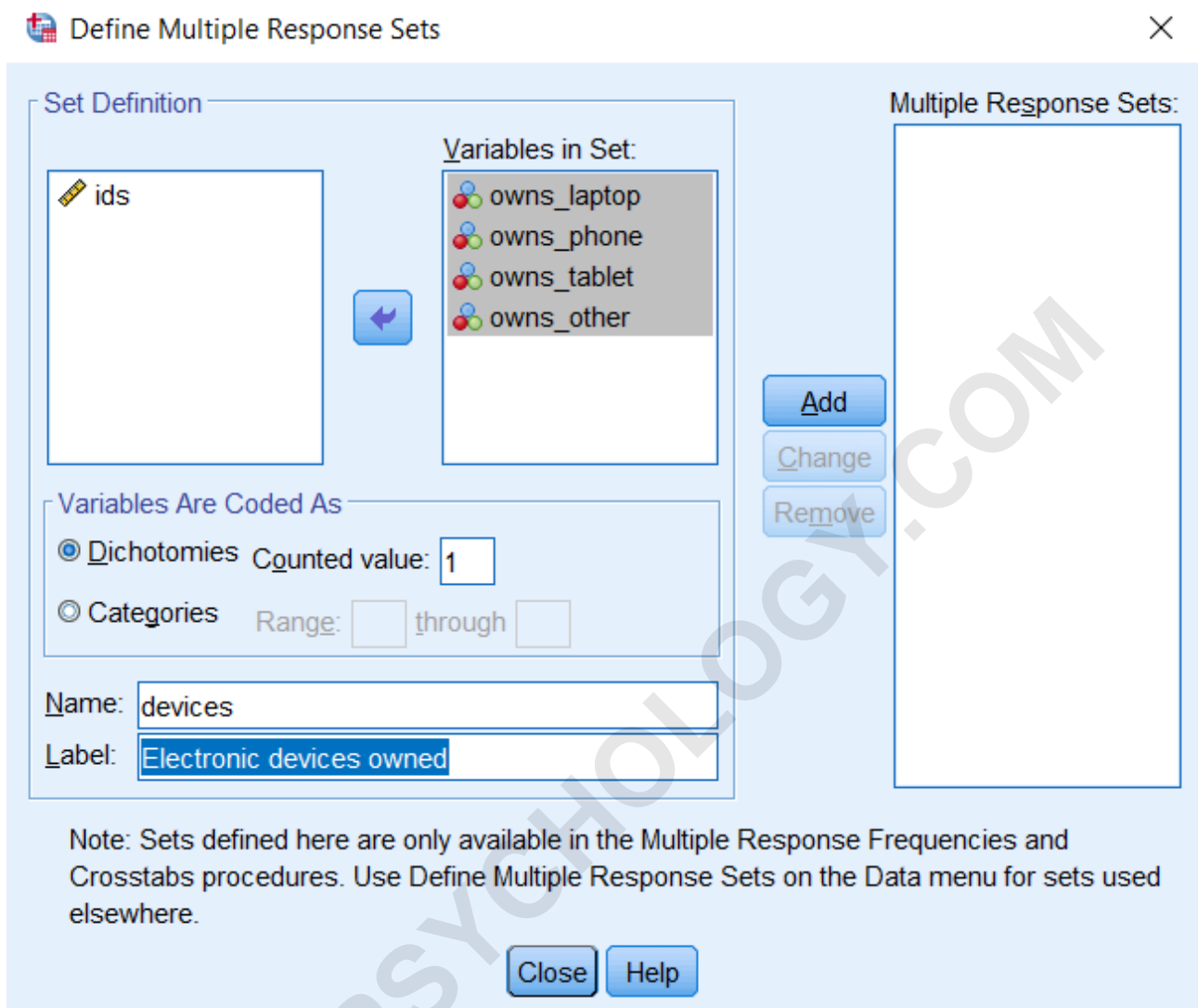
[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.sav](#)

Running the Procedure

Using the Dialog Windows

If using the dialog windows, we must do this in two steps: first, using the Define Multiple Response window, and then using the Multiple Response Frequencies window.

Open the Define Response Set window (**Analyze > Multiple Response > Define Set**). Highlight the four multiple response variables (click variable owns_laptop, then hold down Shift and click variable owns_other) in the left column. Then click the arrow button to move them to the Variables in Set box. In the Variables Coded As section, in the box labeled Counted Value, type 1. In the Name box, type a new name for the set; in this case, we'll use the set name devices. In the Label box, type a descriptive label; in this case, we'll use "Electronic devices owned".



Define Multiple Response Sets [X]

Set Definition

ids

Variables in Set:

- owns_laptop
- owns_phone
- owns_tablet
- owns_other

Variables Are Coded As

Dichotomies Counted value: 1

Categories Range: through

Name: devices

Label: Electronic devices owned

Multiple Response Sets:

Add
Change
Remove

Note: Sets defined here are only available in the Multiple Response Frequencies and Crosstabs procedures. Use Define Multiple Response Sets on the Data menu for sets used elsewhere.

Close Help

Click **Add** to save your new set. When finished, click **Close**.

After clicking Close, nothing will appear to happen; this is normal. To actually create the table, we now run the Multiple Response Frequencies procedure:

Click **Analyze > Multiple Response > Frequencies**. In the left box, double-click on the new variable set, devices, to move it to the right box. Click **OK**.

Using Syntax

Using syntax for multiple response frequency tables is much simpler: the definition of the set and the command to produce the frequency table are done in the same command:

```
MULT RESPONSE GROUPS=$devices 'Electronic devices owned' (owns_laptop owns_phone
owns_tablet
owns_other (1))
```

```
/FREQUENCIES=$devices.
```

Output

Running the above steps or syntax produces the following output:

Multiple Response

Case Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
\$devices ^a	429	98.6%	6	1.4%	435	100.0%

a. Dichotomy group tabulated at value 1.

\$devices Frequencies

		Responses		Percent of Cases
		N	Percent	
Electronic devices owned ^a	Devices owned: Laptop	397	40.1%	92.5%
	Devices owned: Phone	386	39.0%	90.0%
	Devices owned: Tablet	168	17.0%	39.2%
	Devices owned: Other	40	4.0%	9.3%
Total		991	100.0%	231.0%

a. Dichotomy group tabulated at value 1.

The first table, **Case Summary**, counts the number of cases with valid and "true" nonmissing values -- i.e., cases that did not have any of the options checked. We see that only 6 cases did not select any of the answer options. This matches what we saw from the Count Values Within Cases procedure (above).

The second table, **\$devices Frequencies**, is the frequency table of interest. From left to right, the columns of this table show:

First column: The name or label of the multiple response set. Second column: The variable names or variable labels (if assigned) of the variables in the multiple response set. **N**: The number of cases who selected that response option. Notice that these values match the valid values in the frequency tables from the "basic" Frequencies procedure. The number in the Total row is the total number of selections. **Percent**: The proportion of *selections* accounted for by this category. This

column will always sum to 100%. You can confirm the values in this column by dividing the N of that row by the Total N from the last row of the table (991). **Percent of Cases:** The proportion of the cases (i.e., survey respondents) accounted for by this category. This column's sum will be greater than 100%, but the individual proportions can be interpreted as the prevalence of that option among the survey sample. This is often more meaningful than the Percent value. You can confirm the values in this column by dividing the N of that row by the Valid N from the Case Summary table (429).

Interpretation

Using the values of **N** and **Percentage of Cases** from the multiple response frequency table, we can fill in the table from the beginning of this example:

Devices owned	n	% of respondents (n=429)
Laptop	397	92.5%
Phone	386	90.0%
Tablet	168	39.2%
Other	40	9.3%

This table tells us that:

There were 429 students who responded to the question, i.e. selected at least one of the four device type options. The vast majority of the respondents owned a laptop (92.5% or 397/429) The vast majority of the respondents owned a phone (90.0% or 386/429) Less than half of the respondents owned a tablet (39.2% or 168/429) Less than 10% said they owned some other type of electronic device (9.3% or 40/429)

Limitations

We saw that the Multiple Response Frequencies procedure will treat an individual as "missing" (i.e. did not answer the question) if the individual had missing values for all variables in the set. However, our survey question only had four options -- laptop, phone, tablet, and "other". All of these options assume that the respondent owns an electronic device. **If someone does not have an electronic device, the only way they can accurately respond is to not select any choices!** This means that we can't distinguish between people who don't own any electronic devices and people who skipped the question. Given our original research question, this would be especially problematic: if we are interested in knowing the electronic devices that college students own, we need to be certain about what proportion of students do not own any devices, since that could impact students' access to online course materials.

How can we prevent this problem when designing future surveys? One option is to add an answer choice that would specifically accommodate individuals who don't own an electronic device. If we do this, we will need to take an extra step to prevent respondents from giving contradictory answers: for example, we don't want to allow the option for someone to answer "I own a phone and I don't own any electronic devices". Some online survey platforms (such as Qualtrics) allow the survey designer to designate specific answer options as "exclusive". Answers marked as "exclusive" will be "either-or": you can choose any and all of the non-exclusive options, or you can choose the exclusive option, but not both simultaneously.

Remember: a good multiple choice question will have answers that span the full range of possible answers. (This is true for both single-choice and check-all-that-apply question types!) Consider your research question, and use it to guide whether you should include an option like "other", "not applicable", or "none of these".

Example: Creating a Multiple Response Crosstab

Problem Statement

We've gone over how to do frequency tables for multiple response variables; in that example, our concern was counting how common each of the electronic device options were. What if we want to compare differences in device ownership between independent groups, such as men and women? Our desired table of results might look like this:

	Males	Females	Overall			
Devices owned	n	% (nMales=??)	n	% (nFemales=??)	n	% (nTotal=??)
Laptop						
Phone						
Tablet						
Other						

We would like to obtain a crosstab, but as we saw in the previous example, the regular Crosstab procedure does not work the way we would expect when multiple response set variables are involved:

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
	Devices owned: Laptop * Gender	378	86.9%	57	13.1%	435
Devices owned: Phone * Gender	368	84.6%	67	15.4%	435	100.0%
Devices owned: Tablet * Gender	162	37.2%	273	62.8%	435	100.0%
Devices owned: Other * Gender	40	9.2%	395	90.8%	435	100.0%

ARABPSYCHOLOGY.COM

Devices owned: Laptop * Gender Crosstabulation

Count

		Gender		Total
		Male	Female	
Devices owned: Laptop	1	179	199	378
Total		179	199	378

Devices owned: Phone * Gender Crosstabulation

Count

		Gender		Total
		Male	Female	
Devices owned: Phone	1	175	193	368
Total		175	193	368

Devices owned: Tablet * Gender Crosstabulation

Count

		Gender		Total
		Male	Female	
Devices owned: Tablet	1	82	80	162
Total		82	80	162

Devices owned: Other * Gender Crosstabulation

Count

		Gender		Total
		Male	Female	
Devices owned: Other	1	19	21	40
Total		19	21	40

Crosstabs

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
selected * Gender	414	95.2%	21	4.8%	435	100.0%

selected * Gender Crosstabulation

Count

		Gender		Total
		Male	Female	
selected	.00	3	3	6
	1.00	19	12	31
	2.00	103	123	226
	3.00	70	69	139
	4.00	5	7	12
Total		200	214	414

Recall that the Crosstabs procedure can only use cases that have nonmissing values for both variables. Since unselected values are coded as missing values, the Crosstabs procedure drops them from the table entirely. What we really want is a table that will only drop cases if they're missing values for gender or for the multiple response set (i.e., didn't select any of the answer options). These tables alone won't give us the information we need to fill in the table above. We'll need to use the Multiple Response Crosstabs procedure instead.

Sample Dataset for This Example

[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.sav](#)

[Sample Dataset 2019 - Multiple Response Question with Categorical Variable.csv](#)

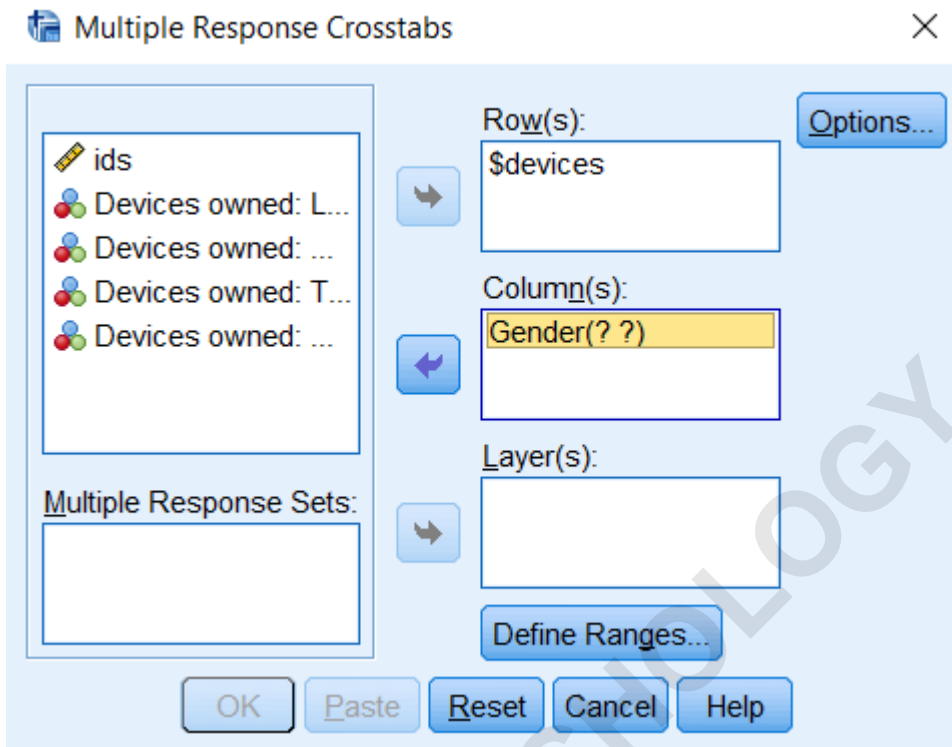
Running the Test

Using the Dialog Windows

If you have not done so already, follow the instructions above to define the multiple response set. Then do the following:

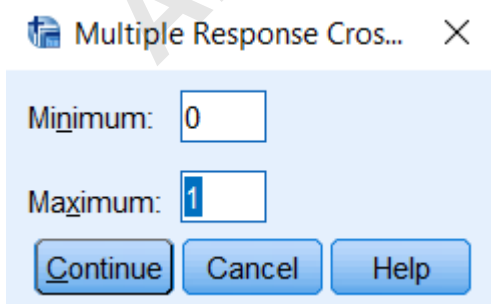
Click **Analyze > Multiple Response > Crosstabs**. In the Multiple Response Sets box, click

variable \$devices. Then click the arrow next to the Rows box. In the first list of variables, click variable Gender. Then click the arrow next to the Columns box. You should see variable Gender appear with two question marks next to it:

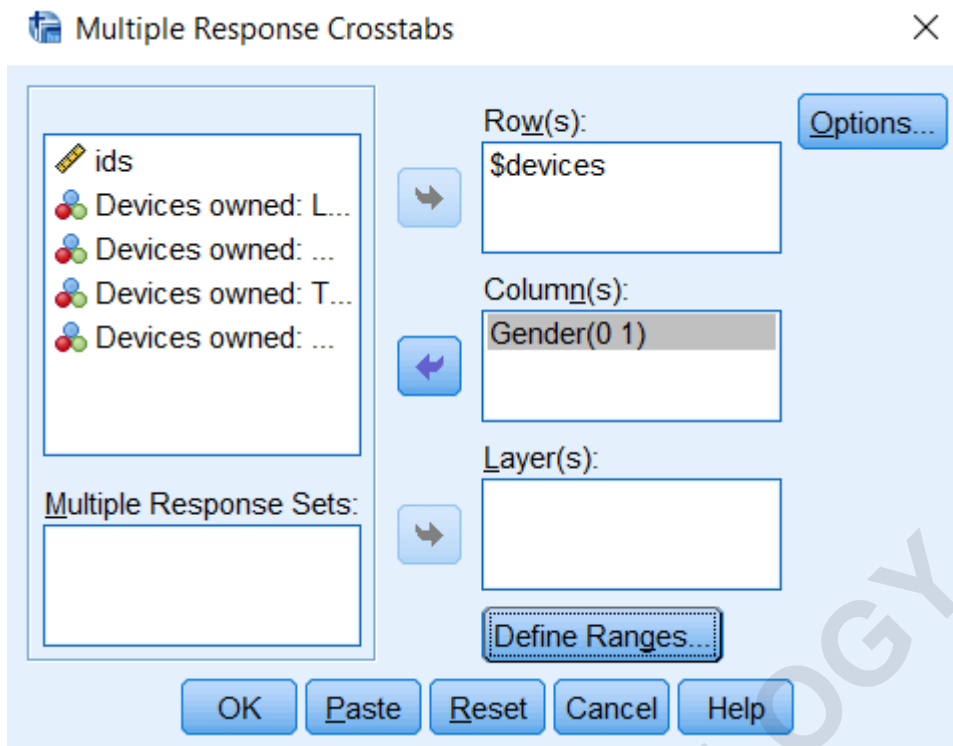


Unlike the normal Crosstabs procedure, we need to specify the range of numeric codes we want to be included in the table. (This must be done for any variable in our crosstab that isn't a multiple response set.) Click **Define Ranges**.

Variable Gender has been coded as Gender=0 for males and Gender=1 for females. In the Minimum box, type 0. In the Maximum box, type 1. Then click **Continue**.



In the Columns box, you should now see our new range appear next to variable Gender.



(Optional) Click **Options**. Under Cell Percentages, select the box for Column, then click **Continue**. Click **OK** to finish.

Using Syntax

Notice that the same procedure, `MULT RESPONSE`, powers both the multiple response frequencies and multiple response crosstabs:

```
MULT RESPONSE GROUPS=$devices 'Electronic devices owned' (owns_laptop owns_phone
owns_tablet
owns_other (1))
/VARIABLES=Gender(0 1)
/TABLES=$devices BY Gender
/BASE=CASES
/CELLS=COLUMN.
```

In this syntax:

`MULT RESPONSE` is the name of the command `GROUPS=` contains the definition of the multiple response group. The definition appears to the right of the equals sign, and has the following components:

`$devices` is the name of the multiple response set; note the dollar sign prefix, which denotes that

this is a multiple response variable. The label for the multiple response set appears in quotation marks. The variables in the multiple response set appear in parentheses. After the name of the last variable (but before the closing parenthesis), we put the number code to count (1) in its own set of parentheses. `/VARIABLES=` specifies any non-multiple-response variables to be used in the tables, and which numeric values (given in parentheses) should be included in the table (in this case, 0 and 1). `/TABLES=` gives the desired structure of the table to the right of the equals sign, using the pattern `rowvar BY colvar BY layervar`. The names of the `rowvar`, `colvar`, and `layervar` variables must match the name of a multiple response set defined in the `GROUPS` statement or a variable name given in the `VARIABLES` statement. `/BASE=CASES` says that, if proportions are computed, use the total number of cases as the denominator. `/CELLS= COLUMN` says to print column percentages in the table.

Output

Multiple Response

Case Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
\$devices*Gender	408	93.8%	27	6.2%	435	100.0%

\$devices*Gender Crosstabulation

		Gender		Total	
		Male	Female		
Electronic devices owned ^a	Devices owned: Laptop	Count	179	199	378
		% within Gender	90.9%	94.3%	
	Devices owned: Phone	Count	175	193	368
		% within Gender	88.8%	91.5%	
	Devices owned: Tablet	Count	82	80	162
		% within Gender	41.6%	37.9%	
	Devices owned: Other	Count	19	21	40
		% within Gender	9.6%	10.0%	
Total	Count	197	211	408	

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

Interpretation

Using the column proportions, we can observe that:

The rate of tablet ownership was slightly higher among males (41.6% of males) than females (37.9% of females). The rate of laptop ownership was approximately four percentage points higher among females than males (94.3% of females versus 90.9% of males).

We can now fill in our table:

	Males		Females		Overall	
Devices owned	n	% (nMales=197)	n	% (nFemales=211)	n	% (nTotal=408)
Laptop	179	90.9%	199	94.3%	378	92.6%
Phone	175	88.8%	193	91.5%	368	90.2%
Tablet	82	41.6%	80	37.9%	162	39.7%
Other	19	9.6%	21	10.0%	40	9.8%

Warning: Do not use the chi-square test of independence on a crosstab containing a multiple response variable. One of the assumptions of the chi-square test of independence is that the responses are uncorrelated with each other. This includes situations where a subject or respondent is counted "twice". Because a respondent can select more than one of the multiple response options, respondents can be counted multiple times, and their responses will be inherently correlated, which violates this critical assumption.

Tutorial Feedback