

What are frequency tables and how are they used in data analysis?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are frequency tables and how are they used in data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149621>

Frequency tables are a type of data representation tool that displays the number of occurrences of each unique value in a dataset. They are used in data analysis to summarize categorical data and provide a visual representation of the frequency distribution of the data. This allows for a quick and easy understanding of the data, helping to identify patterns, trends, and outliers. Frequency tables are commonly used in various fields, such as statistics, market research, and social sciences, to organize and analyze data in a clear and concise manner. They serve as a fundamental tool in data analysis, providing valuable insights and aiding in the decision-making process.

Introduction

When summarizing a *categorical* or *qualitative* variable (nominal/ordinal), we are typically interested in questions like:

How many unique categories were there? How many observations fell into each category? (Counts/frequencies)

Were there any categories with zero observations? Were there any observations with missing responses? What proportion of the observations fell into each category?

What proportion of the non-missing responses fell into each category? What proportion of all responses (missing and non-missing) fell into each category? What is the most frequently occurring category? (Mode)

All of these questions can be answered using a *frequency table*. The most basic type of frequency table contains at least the following:

One row per category of the variable, plus a row for the sum total
A column showing the number of responses in that category
A column showing the proportion of the total observations in that category

In SPSS Statistics, the Frequencies procedure can produce frequency tables, which contain tallies and proportions, as well as two types of graphs appropriate for categorical data: bar charts and pie charts.

Data Requirements

Your data must contain at least one categorical variable that meets all of the following requirements:

The variable must have at least two or more categories (groups), or must be a discrete numeric variable.

The categories may be unordered (nominal) or ordered (ordinal). Each case can be classified into exactly one of the response categories; i.e., a case cannot belong to more than one of the

response categories.

If a case can belong to more than one of the response categories -- for example, responses to a check-all-that-apply survey question -- you should use a Multiple Response Frequency Table instead.

Data Set-Up

Each row can represent one subject, or can represent an observation from a subject. Each column should represent one variable. Variables that will be tabulated using frequency tables should ideally have the following variable properties defined:

Variable Type: The categorical variables in your SPSS dataset can be numeric or string. By default, the rows of the table are arranged in ascending order (for numeric codes) or alphabetically (for string variables).

Ordinal variables may be represented using numbers. Discrete numeric variables should ideally have a limited number of unique values.

Value Labels: If you have entered data using numeric codes that represent specific named categories (especially nominal/unordered categories), you should apply value labels to your variables. This can affect the display of the table.

Missing Value Handling: The frequency table will include sections for Valid (non-missing) and Missing responses. Any values recognized as system-missing or user-missing will appear in the Missing section. If you have more than one user-defined missing value code that appears in the data, those codes are tallied separately in the Missing section of the table. (For example, if you have defined the number code -99 as "Refused response" and -88 to represent "Not asked", you will be able to see how many "Refused" and "Not asked" values there were.)

For numeric variables: Blank values will appear section of the table labeled "Missing". If your dataset also used special number codes to represent missing values (e.g. using), there will be multiple rows in the "Missing" section of the table.

For string variables: Blank strings will NOT automatically be recognized as missing values. If present, they will appear as one of the categories in the "Valid" section; see example below. This can be resolved by using the Automatic Recode procedure to convert the original string variable to a coded numeric variable prior to creating the frequency table.

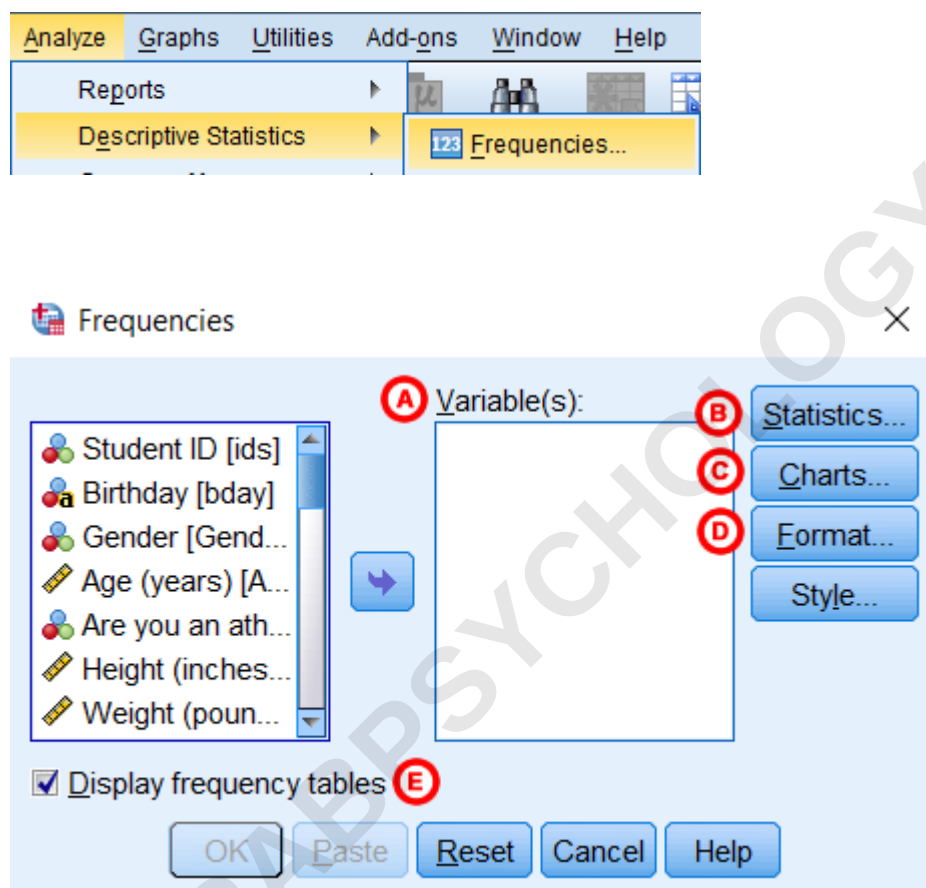
Variable Measurement Levels: The variables' measurement levels should be defined as nominal or ordinal. The Frequencies procedure will still work on variables whose measurement level is set to scale; however, frequency tables should only be used when there are a limited number of

response categories.

Create a Frequency Table in SPSS

In SPSS, the Frequencies procedure can produce summary measures for categorical variables in the form of frequency tables, bar charts, or pie charts.

To run the Frequencies procedure, click **Analyze > Descriptive Statistics > Frequencies**.



A Variable(s): The variables to produce Frequencies output for. To include a variable for analysis, double-click on its name to move it to the Variables box. Moving several variables to this box will create several frequency tables at once.

B Statistics: Opens the Frequencies: Statistics window, which contains various descriptive statistics.

Frequencies: Statistics

Percentile Values

☐ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add

Change

Remove

Central Tendency

☐ Mean

☐ Median

☐ Mode

☐ Sum

☐ Values are group midpoints

Dispersion

☐ Std. deviation ☐ Minimum

☐ Variance ☐ Maximum

☐ Range ☐ S.E. mean

Characterize Posterior Di...

☐ Skewness

☐ Kurtosis

Continue Cancel Help

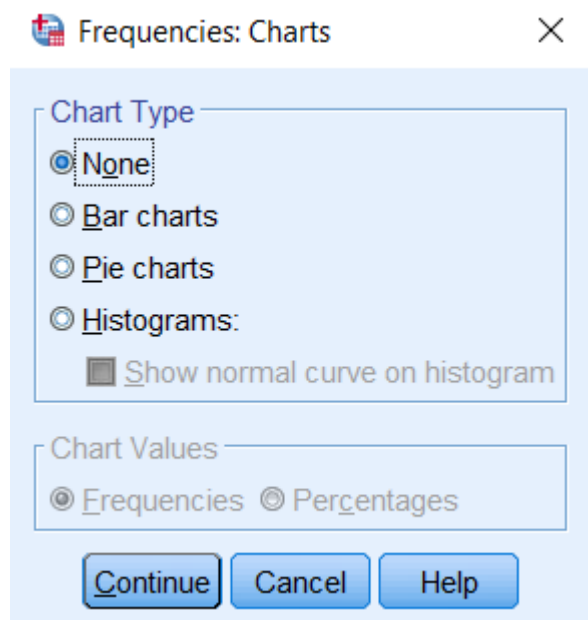
The vast majority of the descriptive statistics available in the Frequencies: Statistics window are never appropriate for nominal variables, and are rarely appropriate for ordinal variables in most situations. There are two exceptions to this:

The **Mode** (which is the most frequent response) has a clear interpretation when applied to most nominal and ordinal categorical variables. The **Values are group midpoints** option can be applied to certain ordinal variables that have been coded in such a way that their value takes on the midpoint of a range. For example, this would be the case if you had measured subjects' ages and had coded anyone between the ages of 20 and 29 as 25, or between the 30 and 39 as 35 (source: [IBM SPSS Statistics Information Center](#)).

If your categorical variables are coded numerically, it is very easy to mis-use measures like the mean and standard deviation. SPSS will compute those statistics if they are requested, regardless of whether or not they are meaningful. It is up to the researcher to determine if these measures are appropriate for their data. In general, you should never use any of these statistics for dichotomous variables or nominal variables, and should only use these statistics with caution for ordinal variables.

CCharts: Opens the Frequencies: Charts window, which contains various graphical options.

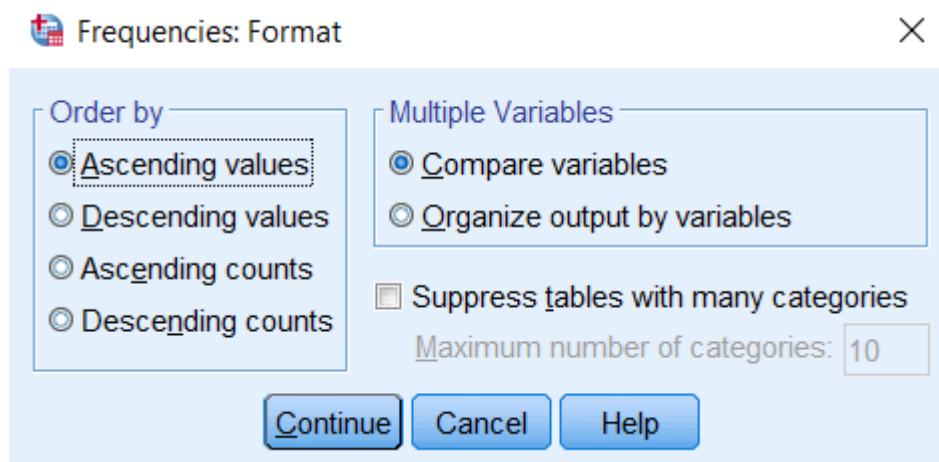
Options include bar charts, pie charts, and histograms. For categorical variables, bar charts and pie charts are appropriate. Histograms should only be used for continuous variables; they should *not* be used for ordinal variables, and should *never* be used with nominal variables.



Bar chart displays the categories on the graph's x-axis, and either the frequencies or the percentages on the y-axis. **Pie chart** depicts the categories of a variable as "slices" of a circular "pie".

Note that the options in the **Chart Values** area apply only to bar charts and pie charts. In particular, these options affect whether the labeling for the pie slices or the y-axis of the bar chart uses counts or percentages. This setting will be greyed out if Histograms is selected.

DFormat: Opens the Frequencies: Format window, which contains options for how to sort and organize the table output.



The **Order by** options affect only categorical variables:

Ascending values arranges the rows of the frequency table in increasing order with respect to the category values: (alphabetically if string, or by numeric code if numeric) **Descending values** arranges the rows of the frequency table in decreasing order with respect to the category values.

Note: If your categorical variable is coded numerically as 0, 1, 2, ..., sorting by ascending or descending value will arrange the rows with respect to the numeric code, **not** with respect to any assigned labels.) **Ascending counts** orders the rows of the frequency table from least frequent (lowest count) to most frequent (highest count). **Descending counts** orders the rows of the frequency table from most frequent (highest count) to least frequent (lowest count).

When working with two or more categorical variables, the **Multiple Variables** options only affects the order of the output. If **Compare variables** is selected, then the frequency tables for all of the variables will appear first, and all of the graphs for the variables will appear after. If **Organize output by variables** is selected, then the frequency table and graph for the first variable will appear together; then the frequency table and graph for the second variable will appear together; etc.

Display frequency tables: When checked, frequency tables will be printed. (This box is checked by default.) If this check box is not checked, no frequency tables will be produced, and the only output will come from supplementary options from **Statistics** or **Charts**. **For categorical variables, you will usually want to leave this box checked.**

Example: Summarizing a Categorical Variable

Using the sample dataset, let's create a frequency table and a corresponding bar chart for the class rank (variable *Rank*), and let's also request the Mode statistic for this variable.

Running the Procedure

Using the Frequencies Dialog Window

Open the Frequencies window (**Analyze > Descriptive Statistics > Frequencies**) and double-click on variable *Rank*. To request the mode statistic, click **Statistics**. Check the box next to **Mode**, then click **Continue**. To turn on the bar chart option, click **Charts**. Select the radio button for **Bar Charts**. Then click **Continue**. When finished, click **OK**.

Using Syntax

```
FREQUENCIES VARIABLES=Rank  
/STATISTICS=MODE  
/BARCHART FREQ  
/ORDER=ANALYSIS.
```

Output

Two tables appear in the output: **Statistics**, which reports the number of missing and nonmissing observations in the dataset, plus any requested statistics; and the frequency table for variable *Rank*. The table title for the frequency table is determined by the variable's label (or the variable name, if a label is not assigned).

Statistics		
Class rank		
N	Valid	406
	Missing	29
Mode		1

Here, the Statistics table shows that there are 406 valid and 29 missing values. It also shows the Mode statistic: here, the mode value is "1", which is the numeric code for the category Freshman. Notice that the Mode statistic isn't displaying the value labels, even though they have been assigned. (For this reason, we recommend not requesting the mode statistic; instead, determine the mode from the frequency table.)

Class rank

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Freshman	147	33.8	36.2	36.2
	Sophomore	96	22.1	23.6	59.9
	Junior	98	22.5	24.1	84.0
	Senior	65	14.9	16.0	100.0
	Total	406	93.3	100.0	
Missing	System	29	6.7		
Total		435	100.0		

Notice how the rows are grouped into "Valid" and "Missing" sections. This grouping allows for easy comparison of missing versus nonmissing observations. Note that "System" missing responses are observations that use SPSS's default symbol -- a period (.) -- for indicating missing values. If a user has assigned special codes for missing values in the Variable View window, those codes would appear here.

The frequency table contains four columns of summary measures:

The **Frequency** column indicates how many observations fell into the given category.

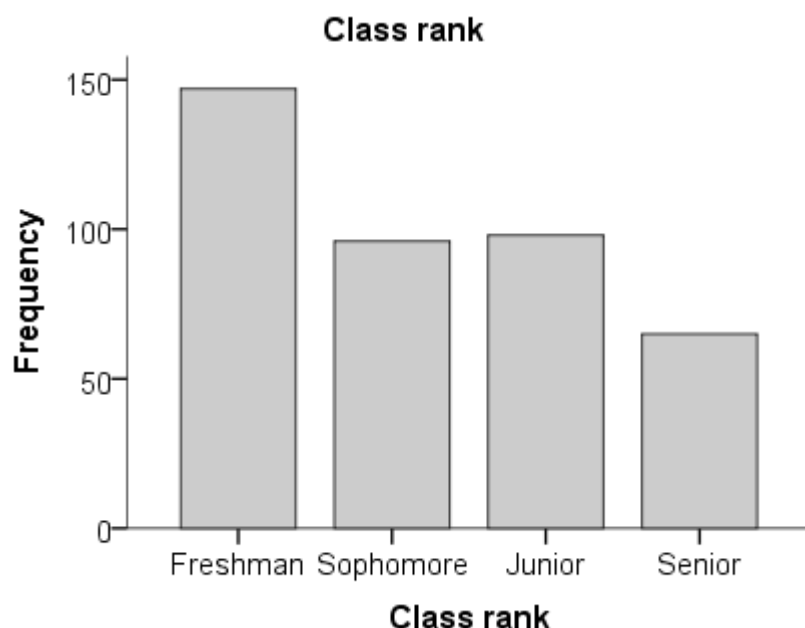
The sample contained a total of 435 students. Of those students, 29 did not specify their class rank. The **Percent** column indicates the percentage of observations in that category out of all observations (both missing and nonmissing). You can verify the proportions for each group by dividing its count in the "frequency" column by the value in the last row of the table (435):

Freshman: $147/435 = 33.8\%$ Sophomore: $96/435 = 22.1\%$ Junior: $98/435 = 22.5\%$ Senior: $65/435 = 14.9\%$ Valid Total: $406/435 = 93.3\%$ Missing: $29/435 = 6.7\%$ The **Valid Percent** column displays the percentage of observations in that category out of the total number of *nonmissing* responses. You can verify the proportions for each group by dividing its count in the "frequency" column by the value of "Total" that appears after the last valid category (406):

Freshman: $147/406 = 36.2\%$ Sophomore: $96/406 = 23.6\%$ Junior: $98/406 = 24.1\%$ Senior: $65/406 = 16.0\%$ The **Cumulative Percent** column is the total percentage of the sample that has been accounted for up to that row; it can be computed by adding all of the numbers in the Valid Percent column above the current row:

Freshman: 36.2% (there are no rows before this one, so the first cumulative percent is identical to the first valid percent) Sophomore: $36.2 + 23.6 = 59.8\%$ Junior: $36.2 + 23.6 + 24.1 = 83.9\%$ Senior: $36.2 + 23.6 + 24.1 + 16.0 = 100\%$

The bar chart appears after the tables.



Here, we can see that:

Freshmen comprised the largest group There were approximately equal numbers of sophomores and juniors Seniors were the smallest group

What if my frequency table has a blank row in it?

What should I do if I create a frequency table in SPSS and one of the rows is blank?

If you are creating a frequency table and notice that the first row has a blank category label, similar to this example:

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	27	6.2	6.2	6.2
In state	314	72.2	72.2	78.4
Out of state	94	21.6	21.6	100.0
Total	435	100.0	100.0	

This issue should not be ignored! **This particular issue affects frequency tables created from string variables that use blanks to denote missing values.** SPSS does not automatically recognize blank (i.e., empty) strings as missing values, so the blank values appear as one of the "Valid" (i.e., non-missing) categories. This affects the calculation of the Valid Percent columns.

When missing values are treated as valid values, it causes the "Valid Percent" columns to be calculated incorrectly. If the blank values were correctly treated as missing values, the valid, non-missing sample size for this table would be **314 + 94 = 408** -- not 435! -- and the valid percent values would change to $314/408 = 76.9\%$ and $94/408 = 23.0\%$. Depending on the number of missing values in your sample, the differences could be even more dramatic.

To fix this problem: To get SPSS to recognize blank strings as missing values, you'll need to run the variable through the Automatic Recode procedure. This procedure takes a string variable and converts it to a new, coded numeric variable with value labels attached. During this process, blank string values are recoded to a special missing value code. To see a worked example, see the [Automatic Recode tutorial](#).

Why are some categories missing from my frequency table and how do I get them to display?

In the sample dataset, variable HowCommute has the following value labels defined: 1=Walk, 2=Bike, 3=Car, 4=Public Transit, and 5=Other. However, if you use the Frequencies procedure to create a frequency table for this variable, you will only see four of the five categories:

How do you commute to campus?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Walk	26	6.0	10.5	10.5
	Bike	15	3.4	6.1	16.6
	Car	193	44.4	78.1	94.7
	Public transit	13	3.0	5.3	100.0
	Total	247	56.8	100.0	
Missing	System	188	43.2		
Total		435	100.0		

Why is one of the categories missing despite it being defined as a value label?

The Frequencies procedure is designed to drop unobserved categories from the frequency table: that is, it will not include categories with counts of 0. Although this can be desirable in some cases, it may be actively problematic or misleading in others. For example, if you create a frequency table of a 5-point Likert item or multiple choice question, readers may interpret the omission of categories as the categories not being included in the design of the survey -- which is very different than the categories being present on the survey but not selected by any respondents.

If you wish to create a frequency table that will include *all* categories with a defined value label even if they have counts of 0, you must use the Custom Tables procedure.

The Custom Tables procedure is included with SPSS Statistics Standard and SPSS Statistics Premium, but is not included in SPSS Statistics Base. If you do not see the Custom Tables procedure in the Analyze menu (**Analyze > Tables > Custom Tables**), it is possible your license did not include the Custom Tables module.

You can check which modules are available to you by opening a new syntax window (**File > New > Syntax**) and executing the following command:

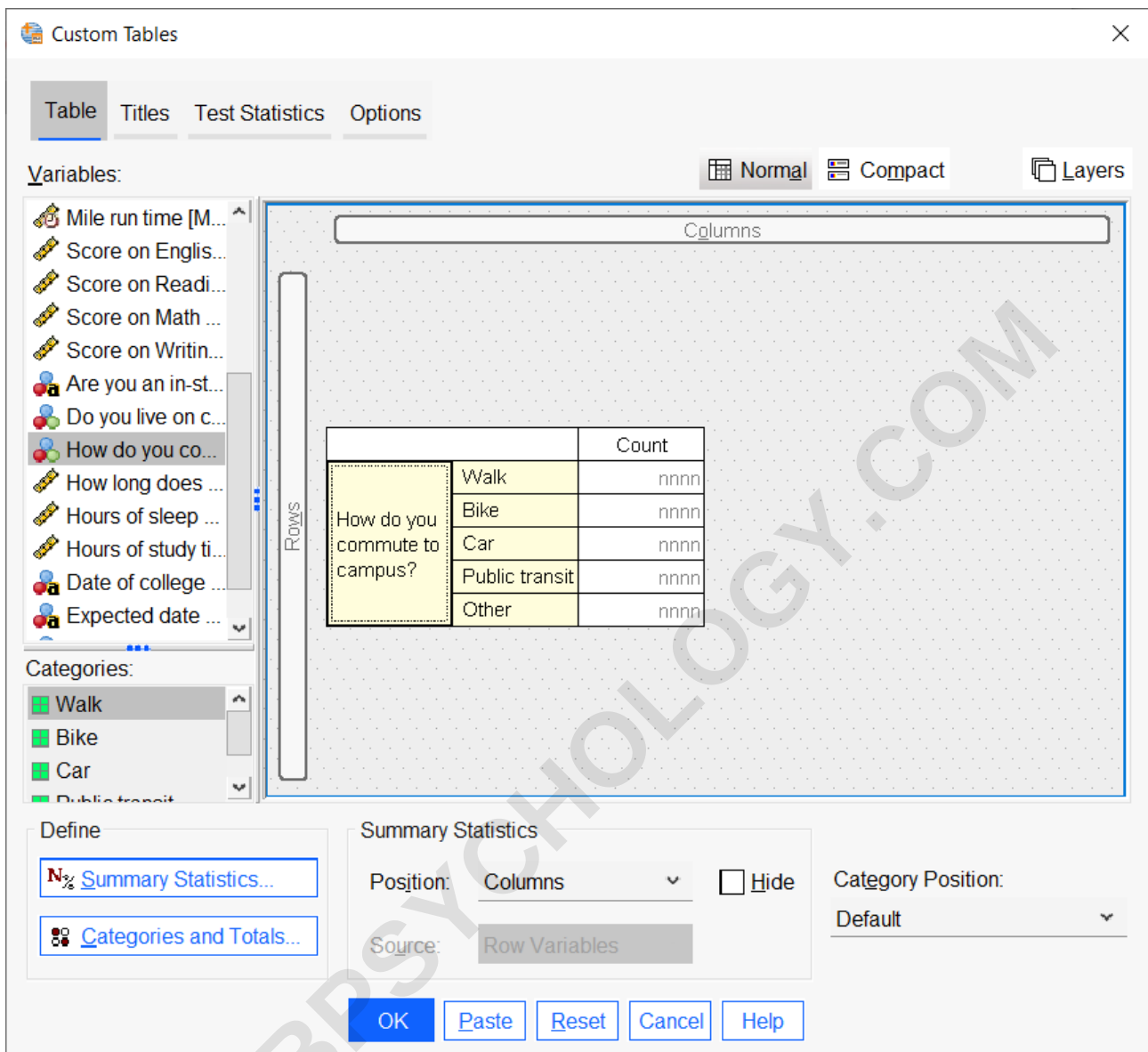
```
SHOW LIC.
```

If the resulting output does not include "IBM SPSS Statistics Custom Tables", then you will not have access to the procedure.

Running the Procedure

Using the Custom Tables Dialog Window

Open the Custom Tables dialog window (**Analyze > Tables > Custom Tables**). Click and drag the HowCommute variable from the list of variables and drop it onto the "Row" box.



(Optional) To add a Total row showing the sum of the counts:
Click the **Categories and Totals** button. Check the Totals box.

Categories and Totals

Selected Variable: How do you commute to campus?

Display

Values

Value(s)	Label
1	Walk
2	Bike
3	Car
4	Public transit
5	Other

Subtotals and Computed Categories

Add Subtotal... Add Category... Edit... Delete

Categories omitted from all subtotals: 0

Sort Categories

By: Value Order: Ascending

Exclude:

Show

☒ Total Label: Total

☐ Missing Values

☒ Empty categories

☒ Other values found when data are scanned

Totals and Subtotals App...

☐ Above categories to which they apply

☒ Below categories to which they apply

Apply Cancel Help

Click **Apply**. (Optional) If you have defined special missing value codes for the variable and want them to appear in the table:

Click the **Categories and Totals** button. Check the Missing Values box. Click **Apply**. (Optional) To add percentages to the table:

Click the **Summary Statistics** button. In the Statistics box, expand the Table Percent category. Then choose one or more of the following options by moving them to the Display box:

Table N %: The denominator of the calculation uses the sum total of the counts in the created table. **Table Valid N %:** The denominator of the calculation uses the sum total of the counts for nonmissing categories in the created table. **Table Total N %:** The denominator of the calculation uses the number of cases in the dataset overall.

Summary Statistics:

Selected Variable: How do you commute to campus?

Statistics:

- ☒ Table Percent
 - Lower CL for Table N %
 - Upper CL for Table N %
 - Standard Error of Table N %
 - Lower CL for Table valid N %
 - Upper CL for Table valid N %
 - Standard Error of Table Valid N %
 - Lower CL for Table Total N %
 - Upper CL for Table Total N %
 - Standard Error of Table Total N %

Display:

Statistics	Label	Format	Decimals
Count	Count	nnnn	0
Table N %	Table N %	nnnn.n%	1
Table Valid N %	Table Valid N %	nnnn.n%	1
Table Total N %	Table Total N %	nnnn.n%	1

☐ Custom Summary Statistics for Totals and Subtotals

Statistics:

- ☒ Count
- ☒ Row Percent
- ☒ Column Percent
- ☒ Layer Percent
- ☒ Layer Column Percent
- ☒ Layer Row Percent
- ☒ Subtable Percent
- ☒ Table Percent
- ☒ Mean, Median, Mode
- ☒ Sum

Display:

Statistics	Label	Format	Decimals
Count	Count	nnnn	0

Confidence Intervals

Level(%):

Apply to Selection Apply to All Close Help

Click **Apply to Selection** to apply your changes to the selected variable, then click **Close** to return to the main dialog window. When finished, click **OK**.

Using Syntax

Default: Counts only (No total row, No percentages)

CTABLES

/VLABELS VARIABLES=HowCommute DISPLAY=LABEL

/TABLE HowCommute

/CATEGORIES VARIABLES=HowCommute ORDER=A KEY=VALUE EMPTY=INCLUDE

/CRITERIA CILEVEL=95.

Counts with total row (No percentages)

CTABLES

/VLABELS VARIABLES=HowCommute DISPLAY=LABEL

/TABLE HowCommute

/CATEGORIES VARIABLES=HowCommute ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=AFTER

```
/CRITERIA CILEVEL=95.
```

Counts and percentages (No total row)

```
CTABLES
/VLABELS VARIABLES=HowCommute DISPLAY=LABEL
/TABLE HowCommute
/CATEGORIES VARIABLES=HowCommute ORDER=A KEY=VALUE EMPTY=INCLUDE
/CRITERIA CILEVEL=95.
```

Counts, percentages, and total row

```
CTABLES
/VLABELS VARIABLES=HowCommute DISPLAY=LABEL
/TABLE HowCommute
/CATEGORIES VARIABLES=HowCommute ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=AFTER
/CRITERIA CILEVEL=95.
```

Output

Default: Counts only (No total row, No percentages)

The default output of Custom Tables includes only the counts. It does not include a total row, nor the number of missing values, nor percentages. However, we see that all 5 categories are present, including the Other category, which is shown with a count of 0:

		Count
How do you commute to campus?	Walk	26
	Bike	15
	Car	193
	Public transit	13
	Other	0

Counts with total row (No percentages)

This table is identical to the previous table, but with the addition of a Total row that includes the sum of the counts in the table. We can verify the total shown in the table by performing the sum manually: $26+15+193+13+0=247$.

		Count
How do you commute to campus?	Walk	26
	Bike	15
	Car	193
	Public transit	13
	Other	0
	Total	247

Counts with percentages and total row

This table is identical to the previous table but with the addition of extra columns containing the percentages we requested.

		Count	Table N %	Table Valid N %	Table Total N %
How do you commute to campus?	Walk	26	10.5%	10.5%	6.0%
	Bike	15	6.1%	6.1%	3.4%
	Car	193	78.1%	78.1%	44.4%
	Public transit	13	5.3%	5.3%	3.0%
	Other	0	0.0%	0.0%	0.0%
	Total	247	100.0%	100.0%	100.0%

In this example, Table N% and Table Valid N% are identical, but this will not always be the case. If your variable includes user-missing values and you have enabled the Missing option, then Table N% will be different than Table Valid N%. (In general, "Table Valid N%" in Custom Tables has the same meaning as the "Valid Percent" column of the Frequencies output: it is the proportion based on the number of valid, nonmissing cases.)

The Table Total N% values are based on the total number of cases in the dataset (valid + missing). Recall that the sample dataset has 435 rows, and we know from the Frequencies procedure that variable HowCommute has 247 valid/observed values and 188 missing values (247 valid + 188 missing = 435 total). We can verify that the Table Total N% values are based on the number of rows by performing the divisions ourselves and rounding to one decimal place:

Walk: $26/435 = 6.0\%$ Bike: $15/435 = 3.4\%$ Car: $193/435 = 44.4\%$ Public Transit: $13/435 = 3.0\%$ Other: $0/435 = 0.0\%$