

How can I count how many cases are missing in a string variable in SPSS?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I count how many cases are missing in a string variable in SPSS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=162041>

SPSS is a statistical software program commonly used for data analysis. When working with string variables, it is important to ensure that all cases have valid data. In order to determine the number of missing cases in a string variable, one can use the "Count Missing Values" function in SPSS. This function will provide a count of how many cases have missing data in the specified string variable, allowing for accurate and efficient data analysis. By utilizing this function, researchers can effectively identify and address any missing data in their string variables.

How can I count how many cases are missing in a string variable? | SPSS FAQ

There are at least two ways to find out how many missing cases there are in a string variable. The first way is to use the missing values command to define a missing value for the variable. The second way is to create a new variable which is zero if the value is not missing and one if it is missing. To do that, you will need to use the length and the trim functions with the compute command.

Consider the data set below. We have two string variables,

fname (first name) and lname (last name). Because we have given

fname a length of five and lname a length of eight, both variables are short string variables. (Note that if we had just typed (A), the length of the string variable would

have been one, not the length of the first case, as it would have been in other statistical packages, such as SAS.)

```
data list list / id * fname (A5) lname (A8) age.
```

```
begin data
```

```
1 "Beth" "Jones" .
```

```
2 "Bob" "Jensen" 23
```

```
3 " " "Andersen" 25
```

```
4 "Andy" "Smith" 26
```

```
5 "Al" "Peterson" 21
```

```
6 "Ann" "Glenn" 22
```

```
7 "Pete" " " 29
```

```
8 "Pam" "Wright" 21
```

```
9 " " "Brown" 29
```

```
end data.
```

Notice that there are two missing values for fname, and one missing value each for lname and age.

If you

run this code and look at the SPSS data editor, you will see that the cells of the missing names are empty, but the

cell of missing value of age has a period in it. You will

also notice that while SPSS issued an error message regarding the missing value for age, it did not issue an error message for any of the missing names. This is because blanks (i.e., a null string) is a valid value for a string variable. Now let's look at the frequencies of each variable. We know that there should be seven valid cases and two missing for fname, and eight valid cases and one missing for both lname and age.

freq var = fname lname age.

	FNAME	LNAME	AGE	
N	Valid	9	9	8
Missing	0	0	1	

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid		2	22.2	22.2	22.2
Al	1	11.1	11.1	33.3	
Andy	1	11.1	11.1	44.4	
Ann	1	11.1	11.1	55.6	
Beth	1	11.1	11.1	66.7	
Bob	1	11.1	11.1	77.8	
Pam	1	11.1	11.1	88.9	
Pete	1	11.1	11.1	100.0	

Total	9	100.0	100.0		
--------------	---	-------	-------	--	--

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid		1	11.1	11.1	11.1
Andersen	1	11.1	11.1	22.2	
Brown	1	11.1	11.1	33.3	
Glenn	1	11.1	11.1	44.4	
Jensen	1	11.1	11.1	55.6	
Jones	1	11.1	11.1	66.7	
Peterson	1	11.1	11.1	77.8	
Smith	1	11.1	11.1	88.9	
Wright	1	11.1	11.1	100.0	
Total	9	100.0	100.0		

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	21.00	2	22.2	25.0	25.0
22.00	1	11.1	12.5	37.5	
23.00	1	11.1	12.5	50.0	
25.00	1	11.1	12.5	62.5	
26.00	1	11.1	12.5	75.0	
29.00	2	22.2	25.0	100.0	
Total	8	88.9	100.0		
Missing	System	1	11.1		
Total	9	100.0			

However, this is not what we see. Although age as one missing value, neither fname nor lname have missing values. Unlike missing values for

numeric variables, missing values for string variables are not assigned a period (.). Rather, they are left blank and SPSS does not consider them to be missing. To indicate a missing value in a string variable, you need to use the missing values command and assign a "value" to missing cases. This "value" can be one or more blanks, or a numeric code such as 9999. You can only define missing values for string variables whose length is eight or less (what SPSS calls "short" string variables). It is important to note that there are no system missing values for either short nor long string variables. You can assign different missing values to different variables within the same missing values command, as shown below.

missing values fname lname (" ").

Now let's look at the frequencies.

freq var = fname lname age.

	FNAME	LNAME	AGE	
N	Valid	7	8	8

Missing	2	1	1	
----------------	---	---	---	--

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	AI	1	11.1	14.3	14.3
Andy	1	11.1	14.3	28.6	
Ann	1	11.1	14.3	42.9	
Beth	1	11.1	14.3	57.1	
Bob	1	11.1	14.3	71.4	
Pam	1	11.1	14.3	85.7	
Pete	1	11.1	14.3	100.0	
Total	7	77.8	100.0		
Missing		2	22.2		
Total	9	100.0			

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	Andersen	1	11.1	12.5	12.5
Brown	1	11.1	12.5	25.0	
Glenn	1	11.1	12.5	37.5	
Jensen	1	11.1	12.5	50.0	
Jones	1	11.1	12.5	62.5	
Peterson	1	11.1	12.5	75.0	
Smith	1	11.1	12.5	87.5	
Wright	1	11.1	12.5	100.0	
Total	8	88.9	100.0		
Missing		1	11.1		
Total	9	100.0			

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	21.00	2	22.2	25.0	25.0
22.00	1	11.1	12.5	37.5	

23.00	1	11.1	12.5	50.0	
25.00	1	11.1	12.5	62.5	
26.00	1	11.1	12.5	75.0	
29.00	2	22.2	25.0	100.0	
Total	8	88.9	100.0		
Missing	System	1	11.1		
Total	9	100.0			

Now the frequencies are as we would expect them to be.

You can also use the display dictionary command to see that the missing values have been properly assigned.

display dictionary.

List of variables on the working file

Name Position

ID 1

Measurement Level: Scale

Column Width: 8 Alignment: Right

Print Format: F8.2

Write Format: F8.2

FNAME 2

Measurement Level: Nominal

Column Width: 8 Alignment: Left

Print Format: A5

Write Format: A5

Missing Values: "

LNAME 3

Measurement Level: Nominal

Column Width: 8 Alignment: Left

Print Format: A8

Write Format: A8

Missing Values: "

AGE 4

Measurement Level: Scale

Column Width: 8 Alignment: Right

Print Format: F8.2

Write Format: F8.2

Note that we did not assign any missing values to id or age; therefore, none are shown under those variables.

You can also use the missing values command to

delete previously declared missing values. To do this, do not type anything in the parentheses after the variable

listed on the missing values command. If you use the SPSS keyword

all instead of a variable or a list of

variables, you will delete all user-defined missing values for all variables, both string and numeric.

The second way to determine the number of missing values for a string variable is to create a new variable that has a value of one if the cell in the original variable is not empty (i.e., there is a character of some sort in there,)

and one if it is empty. Next, the describe command is used to sum up the number of ones (i.e., the number of missing

values). For our example, we will create a new variable for each string variable in our data set.

The variable

missf will indicate the missing values for fname, and missl will indicate the

missing values for lname. We will use two functions with the compute command to create our new variables.

The

rtrim function trims the blank spaces from the right of the variable. If there is nothing but spaces, it trims the length to zero. The length function determines length of the value. Next, the expression is evaluated for each case. If the

length is zero, the expression is true and a one is placed in the new variable. If the expression is false, then

a zero is placed in the new variable.

compute missf = (length(rtrim(fname)) = 0).

compute missl = (length(rtrim(lname)) = 0).

execute.

desc var = missf missl

/statistics = sum.

	N	Sum
MISSF	9	2.00
MISSL	9	1.00
Valid N (listwise)	9	

list.

ID FNAME LNAME AGE MISSF MISSL

1.00 Beth Jones . .00 .00
2.00 Bob Jensen 23.00 .00 .00
3.00 Andersen 25.00 1.00 .00
4.00 Andy Smith 26.00 .00 .00
5.00 Al Peterson 21.00 .00 .00
6.00 Ann Glenn 22.00 .00 .00
7.00 Pete 29.00 .00 1.00
8.00 Pam Wright 21.00 .00 .00
9.00 Brown 29.00 1.00 .00

Number of cases read: 9 Number of cases listed: 9

For similar pages, please see How can I count the number of missing values and pattern of missing values in a character variable? and How can I count the number of missing values in a character variable?