

How can PROC CONTENTS be used to summarize the contents of a dataset in SAS?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How can PROC CONTENTS be used to summarize the contents of a dataset in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=150381>

PROC CONTENTS is a SAS procedure that can be used to summarize the contents of a dataset. It provides a comprehensive list of the variables, their data types, lengths, and labels, as well as the number of observations and other important statistical information. This procedure is useful for gaining a quick overview of a dataset and identifying any potential issues or discrepancies. Additionally, it allows for easy navigation and understanding of the dataset's structure, making it a valuable tool for data analysis and management in SAS.

Summarizing Data with PROC CONTENTS

The CONTENTS procedure generates summary information about the contents of a dataset, including:

The variables' names, types, and attributes (including formats, informats, and labels) How many observations are in the dataset How many variables are in the dataset When the dataset was created

This procedure is especially useful if you have imported your data from a file and want to check that your variables have been read correctly, and have the appropriate variable type and format. (For example, you may wish to check that none of your character variables have been truncated, and that your date variables have not been misread.) The basic syntax of PROC CONTENTS is:

```
PROC CONTENTS DATA=sample;  
RUN;
```

As with all SAS procedures, the DATA command (which specifies the name of the dataset) is optional, but recommended. If you do not specify a dataset, SAS will use the most recently created dataset by default.

Note that PROC CONTENTS will list the variables in alphabetical order, rather than the order they appear in the dataset. You can change this by adding the ORDER=VARNUM option to the PROC CONTENTS statement:

```
PROC CONTENTS DATA=sample ORDER=varnum;  
RUN;
```

The screenshot below shows the output of PROC CONTENTS on the sample data file. Key elements are labeled and described below the screenshot.

The SAS System

14:52 Wednesday, May 18, 2016 13

The CONTENTS Procedure

Data Set Name	MYLIB.SAMPLE	(A) Observations	435
Member Type	DATA	(B) Variables	23
Engine	V9	Indexes	0
Created	(C) Wednesday, May 18, 2016 03:06:26 PM	Observation Length	248
Last Modified	Wednesday, May 18, 2016 03:06:26 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	16384
Number of Data Set Pages	7
First Data Page	1
Max Obs per Page	65
Obs in First Data Page	52
Number of Data Set Repairs	0
Filename	E:\Tutorial Sample Data\sample.sas7bdat
Release Created	9.0301M2
Host Created	W32_7PRO

(D) Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
8	Athlete	Num	8	ATHLETECODE.		Athlete
21	CommuteTime	Num	8			Time to commute to campus (minutes)
14	English	Num	8			English
7	Gender	Num	8	GENDERCODE.		Gender
9	Height	Num	8			Height
20	HowCommute	Num	8			HowCommute
19	LiveOnCampus	Num	8			Lives on campus
6	Major	Char	58	\$58.	\$58.	Major
16	Math	Num	8			Math
13	MileMinDur	Num	8	TIME8.	TIME8.	Mile run time (duration)
5	Rank	Num	8	RANKCODE.		Class rank
15	Reading	Num	8			Reading
22	SleepTime	Num	8			Hours of sleep per night
11	Smoking	Num	8	SMOKERCODE.		Smoking

A The number of observations (or rows) in the dataset. Here, the sample dataset contains 435 observations.

B The number of variables (or columns) in the dataset. Here, the sample dataset contains 23 variables.

C The date and time that the dataset was created and last modified.

D This part of the output lists the dataset's variables and their attributes.

(#): The original order of the variable in the columns of the dataset. (PROC CONTENTS prints the variables in alphabetical order with respect to name, instead of in the order that they appear in the dataset.) **Type**: Whether the variable is numeric (Num) or character (Char). **Len**: Short for "Length"; represents the width of the variable. **Format**: The assigned format that will be used when the values of the variable are printed in the Output window. **Informat**: The original format of the variable when it was read into SAS. **Label**: The assigned variable label that will be used when the

name of the variable is printed in the Output window. If your variables do not have labels, this column will be identical to the Variable column.

ARABPSYCHOLOGY.COM