How can I create an empty dataset in SAS?

Authored by

stats writer

June 26, 2024

To create an empty dataset in SAS, you can use the "DATA" statement followed by the desired name for your dataset. After the name, include a semicolon to indicate the end of the statement. This will create a new, empty dataset with no variables or observations. To add variables, you can use the "INPUT" statement followed by the desired variable names and their data types. To add observations, you can use the "SET" statement followed by the name of an existing dataset. This will populate your empty dataset with the same variables and observations as the existing dataset. Alternatively, you can use the "OUTPUT" statement to create a new dataset and specify the desired variables and their data types. This will also create an empty dataset with the specified variables. Overall, creating an empty dataset in SAS involves using specific statements to either add variables or copy existing datasets, depending on your needs.

Create an Empty Dataset in SAS

There are two common ways to create an empty dataset in SAS:

Method 1: Create Empty Dataset from Scratch

```
data empty_data;
attrib
var1 length=8 format=best12. label="var1"
var2 length=$30 format=$30. label="var2"
var3 length=8 format=best12. label="var3"
stop;
run;
```

Method 2: Create Empty Dataset from Existing Dataset

```
data empty_data;
```

```
set existing_data;
stop;
run;
```

In both methods, the stop statement prevents SAS from actually processing any rows.

This results in an empty dataset with variable names but no rows.

The following examples show how to use each method in practice.

Example 1: Create Empty Dataset from Scratch

We can use the following code to create an empty dataset called empty_data that contains four variables:

```
/*create empty dataset*/
data empty_data;
attrib
employee_ID length=8 format=best12. label="Employee ID"
employee_Name length=$30 format=$30. label="Employee Name"
sales length=8 format=best12. label="Sales"
```

```
sales_date length=8 format=date9. label="Sales Date";
stop;
run;
```

We can then use proc contents to view the contents of the dataset:

```
/*view contents of dataset*/
proc contentsdata=empty_data;
```

The CONTENTS Procedure

Data Set Name	WORK.EMPTY_DATA	Observations	0
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	11/17/2022 10:30:43	Observation Length	56
Last Modified	11/17/2022 10:30:43	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

From the output we can see that the dataset has four variables but zero observations, i.e. zero rows.

At the bottom of the output we can also see the names of the four variables we created:

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
1	employee_ID	Num	8	BEST12.	Employee ID
2	employee_Name	Char	30	$30.	Employee Name
3	sales	Num	8	BEST12.	Sales
4	sales_date	Num	8	DATE9.	Sales Date

Example 2: Create Empty Dataset from Existing Dataset

We can use the following code to create an empty dataset called empty_data that is generated from an existing dataset called , which is a dataset built into SAS:

```
/*create empty dataset from existing dataset*/
data empty_dat;
set sashelp.Comet;
stop;
run;
```

We can then use proc contents to view the contents of the dataset:

```
/*view contents of dataset*/
proc contents data=empty_data;
```

The CONTENTS Procedure

Data Set Name	WORK.EMPTY_DAT	Observations	0
Member Type	DATA	Variables	4
Engine	V9	Indexes	0
Created	11/17/2022 10:41:24	Observation Length	32
Last Modified	11/17/2022 10:41:24	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

From the output we can see that the dataset has four variables but zero observations.

At the bottom of the output we can also see the names of the four variables created from the existing dataset:

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Label
1	Dose	Num	8	1,2 Dimethylhydrazine dihydrochloride Dose Level
4	Length	Num	8	Tail Length of the Comet
2	Rat	Num	8	Rat Index
3	Sample	Num	8	Slide Index of Grouped Cells from a Rat

The following tutorials explain how to perform other common tasks in SAS: