

How can I use the SUBSTR function in SAS?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I use the SUBSTR function in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=164324>

The SUBSTR function in SAS is a powerful tool that allows users to extract a portion of a string or character variable based on specified starting and ending positions. This function can be used in various scenarios such as data manipulation, text processing, and data cleaning. To use the SUBSTR function, the user must provide the name of the variable, the starting position, and the desired length of the extracted substring. This function can also be combined with other functions to perform more complex tasks. For example, it can be used with the LENGTH function to extract a specific number of characters from the end of a string. Some examples of using the SUBSTR function in SAS include extracting the first name from a full name variable, removing prefixes or suffixes from a text variable, and converting a numerical string to a numeric variable. Overall, the SUBSTR function is a valuable tool for manipulating and extracting data in SAS programming.

Use the SUBSTR Function in SAS (With Examples)

You can use the SUBSTR function in SAS to extract a portion of a string.

This function uses the following basic syntax:

SUBSTR(Source, Position, N)

where:

Source: The string to analyze

Position: The starting position to read

N: The number of characters to read

Here are the four most common ways to use this function:

Method 1: Extract First N Characters from String

```
data new_data;  
set original_data;  
first_four = substr(string_variable, 1, 4);  
run;
```

Method 2: Extract Characters in Specific Position Range from String

```
data new_data;  
set original_data;  
two_through_five = substr(string_variable, 2, 4);  
run;
```

Method 3: Extract Last N Characters from String

```
data new_data;  
set original_data;  
last_three = substr(string_variable,  
length(string_variable)-2, 3);  
run;
```

Method 4: Create New Variable if Characters Exist in String

```
data new_data;  
set original_data;  
ifsubstr(string_variable, 1, 4) = 'some_string' then  
new_var = 'Yes';  
else new_var = 'No';  
run;
```

The following examples show how to use each method with the following dataset in SAS:

```
/*create dataset*/  
data original_data;  
input team $1-10;  
datalines;  
Warriors  
Wizards  
Rockets  
Celtics  
Thunder  
;  
run;
```

```
/*view dataset*/  
proc printdata=original_data;
```

Obs	team
1	Warriors
2	Wizards
3	Rockets
4	Celtics
5	Thunder

Example 1: Extract First N Characters from String

The following code shows how to extract the first 4 characters from the team variable:

```
/*create new dataset*/  
data new_data;  
set original_data;  
first_four = substr(team, 1, 4);  
run;  
  
/*view new dataset*/proc printdata=new_data;
```

Obs	team	first_four
1	Warriors	Warr
2	Wizards	Wiza
3	Rockets	Rock
4	Celtics	Celt
5	Thunder	Thun

Notice that the first_four variable contains the first four characters of the team variable.

Example 2: Extract Characters in Specific Position Range from String

The following code shows how to extract the characters in positions 2 through 5 from the team variable:

```
/*create new dataset*/  
data new_data;  
set original_data;  
two_through_five = substr(team, 2, 4);  
run;  
  
/*view new dataset*/proc printdata=new_data;
```

Obs	team	two_through_five
1	Warriors	arri
2	Wizards	izar
3	Rockets	ocke
4	Celtics	elti
5	Thunder	hund

Example 3: Extract Last N Characters from String

The following code shows how to extract the last 3 characters from the team variable:

```
/*create new dataset*/  
data new_data;  
set original_data;  
last_three = substr(team, length(team)-2, 3);  
run;  
  
/*view new dataset*/proc printdata=new_data;
```

Obs	team	last_three
1	Warriors	ors
2	Wizards	rds
3	Rockets	ets
4	Celtics	ics
5	Thunder	der

Example 4: Create New Variable if Characters Exist in String

The following code shows how to create a new variable called **W_Team** that takes a value of 'yes' if the first character in the team name is 'W' or a value of 'no' if the first character is not a 'W.'

```
/*create new dataset*/  
data new_data;  
set original_data;  
ifsubstr(team, 1, 1) = 'W' then W_Team = 'Yes';
```

```
else W_Team = 'No';  
run;
```

```
/*view new dataset*/proc printdata=new_data;
```

Obs	team	W_Team
1	Warriors	Yes
2	Wizards	Yes
3	Rockets	No
4	Celtics	No
5	Thunder	No

Additional Resources

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