

How do I extract a substring from the right using the SAS programming language?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I extract a substring from the right using the SAS programming language?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=147953>

The process of extracting a substring from the right using the SAS programming language involves using the SUBSTR function and specifying a negative starting position to indicate the number of characters from the end of the string. This allows for the extraction of a specific portion of the string, starting from the right, for further manipulation or analysis within the SAS program. This technique is useful for handling data that follows a consistent format, such as dates or unique identifiers, and can be easily implemented in SAS code to enhance data processing and analysis.

SAS: Extract Substring from Right

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

The value for the Position argument specifies the starting position *from the left of the string*.

To instead specify a starting position *from the right of the string*, you can use the following syntax:

```
data new_data;  
set original_data;
```

```
last_three = substr(team, length(team)-2, 3);  
run;
```

This particular example creates a new variable called `last_three` that extracts the last three characters from the right of the string variable called `team`.

The following example shows how to use this syntax in practice.

Example: Extract Substring from Right in SAS

Suppose we have the following dataset in SAS that contains information about various basketball teams:

```
/*create dataset*/  
data original_data;  
input team $ points;  
datalines;  
Mavericks 104  
Thunder 99  
Rockets 116  
Spurs 98  
Pistons 99  
Pelicans 105
```

Warriors 119

Blazers 113

Nuggets 100

Kings 123

;

run;

/*view dataset*/

proc printdata=original_data;

Obs	team	points
1	Maverick	104
2	Thunder	99
3	Rockets	116
4	Spurs	98
5	Pistons	99
6	Pelicans	105
7	Warriors	119
8	Blazers	113
9	Nuggets	100
10	Kings	123

We can use the following code 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	points	last_three
1	Maverick	104	ick
2	Thunder	99	der
3	Rockets	116	ets
4	Spurs	98	urs
5	Pistons	99	ons
6	Pelicans	105	ans
7	Warriors	119	ors
8	Blazers	113	ers
9	Nuggets	100	ets
10	Kings	123	ngs

We can also easily change the values in the SUBSTR function to extract a different number of characters from the right side of the string.

For example, we can use the following syntax to extract the last five characters from the right:

```
/*create new dataset*/  
data new_data;
```

```
set original_data;  
last_five = substr(team, length(team)-4, 5);  
run;  
  
/*view new dataset*/proc printdata=new_data;
```

Obs	team	points	last_five
1	Maverick	104	erick
2	Thunder	99	under
3	Rockets	116	ckets
4	Spurs	98	Spurs
5	Pistons	99	stons
6	Pelicans	105	icans
7	Warriors	119	riors
8	Blazers	113	azers
9	Nuggets	100	ggets
10	Kings	123	Kings

Notice that the column called last_five contains the last five characters from the team column.

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