

# How do I extract numbers from a string in SAS?

Authored by  
**stats writer**

June 26, 2024

## RECOMMENDED CITATION

stats writer (2024). *How do I extract numbers from a string in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154217>

SAS is a programming language commonly used for data analysis and statistical modeling. In order to extract numbers from a string in SAS, you can use a combination of the SCAN and INPUT functions. The SCAN function allows you to extract specific words or characters from a string, while the INPUT function converts those extracted values into numeric format. By using these functions together, you can effectively retrieve and manipulate numerical data from a string in SAS. This process is particularly useful when working with data sets that contain both textual and numerical information.

## Extract Numbers from String in SAS

**The easiest way to extract numbers from a string in SAS is to use the function with the 'A' modifier.**

**This function uses the following basic syntax:**

```
data new_data;  
set original_data;  
numbers_only = compress(some_string, ", 'A');  
run;
```

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

### Example: Extract Numbers from String in SAS

**Suppose we have the following dataset in SAS that shows the names of various college courses:**

```
/*create dataset*/
```

```
data original_data;  
input course $12.;  
datalines;  
Stats101  
Economics203  
Business201  
Botany411  
Calculus101  
English201  
Chemistry402  
Physics102  
;  
run;  
  
/*view dataset*/  
proc printdata=original_data;
```

| Obs | course       |
|-----|--------------|
| 1   | Stats101     |
| 2   | Economics203 |
| 3   | Business201  |
| 4   | Botany411    |
| 5   | Calculus101  |
| 6   | English201   |
| 7   | Chemistry402 |
| 8   | Physics102   |

We can use the following code to extract only the numbers from each course name:

```
/*extract numbers from course column*/  
data new_data;  
set original_data;  
course_number_only = compress(course, ", 'A');  
run;  
  
/*view results*/  
proc printdata=new_data;
```

| Obs | course       | course_number_only |
|-----|--------------|--------------------|
| 1   | Stats101     | 101                |
| 2   | Economics203 | 203                |
| 3   | Business201  | 201                |
| 4   | Botany411    | 411                |
| 5   | Calculus101  | 101                |
| 6   | English201   | 201                |
| 7   | Chemistry402 | 402                |
| 8   | Physics102   | 102                |

Notice that the new column called `course_number_only` contains only the numbers from the strings in the `course` column.

If you would instead like to only extract the characters

in each string, you can use the **COMPRESS** function with the 'd' modifier instead:

```
/*extract characters from course column*/  
data new_data;  
set original_data;  
course_characters_only = compress(course, "", 'd');  
run;  
  
/*view results*/  
proc printdata=new_data;
```

| Obs | course       | course_characters_only |
|-----|--------------|------------------------|
| 1   | Stats101     | Stats                  |
| 2   | Economics203 | Economics              |
| 3   | Business201  | Business               |
| 4   | Botany411    | Botany                 |
| 5   | Calculus101  | Calculus               |
| 6   | English201   | English                |
| 7   | Chemistry402 | Chemistry              |
| 8   | Physics102   | Physics                |

**Notice that the new column called `course_characters_only` contains only the numbers from the strings in the `course` column.**

**Note: You can find a complete list of modifiers for the COMPRESS function on this .**

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

ARABPSYCHOLOGY.COM