SUBSTR Function

SCAN versus SUBSTR

The SUBSTR function is similar to the SCAN function. Let's briefly compare the two. Both the SCAN and SUBSTR functions can extract a substring from a character value:

- SCAN extracts words within a value that is marked by delimiters.
- SUBSTR extracts a portion of a value by starting at a specified location.

The SUBSTR function is best used when you know the exact position of the string that you want to extract from the character value. The string does not need to be marked by delimiters.

The SUBSTR function can be used to

- extract a portion of a character value
- replace the contents of a character value.

Extract a portion of a character value

General form, SUBSTR function:

SUBSTR (argument, position, n)

Example: MiddleInitial = substr (middlename,1,1)

where

- argument specifies the character variable or expression to scan.
- position is the character position to start from.
- n specifies the number of characters to extract. If n is omitted, all remaining characters are included in the substring.

Replacing Text Using SUBSTR

There is a second use for the SUBSTR function. This function can also be used to replace the contents of a character variable.

The syntax of the SUBSTR function, when used to replace a variable's values, is identical to the syntax for extracting a substring.

SUBSTR(argument,position,n)

Example: substr(test,4,2)='92';

S7381K2 → S7392K2

S7381K7 → S7392K7

However, in this case,

- the first argument specifies the character variable whose values are to be modified.
- the second argument specifies the position at which the replacement is to begin.
- the third argument specifies the number of characters to replace. If n is omitted, all remaining characters are replaced.

Positioning the SUBSTR Function

SAS uses the SUBSTR function to extract a substring or to modify a variable's values, depending on the position of the function in the assignment statement.

When the function is on the *right side* of an assignment statement, the function returns the requested string.

```
MiddleInitial=substr(middlename,1,1);
```

But if you place the SUBSTR function on the *left side* of an assignment statement, the function is used to modify variable values.

```
substr(region,1,3)='NNW';
```