How can I create dummy variables in SAS with an example?

Authored by

stats writer

July 1, 2024

Creating dummy variables in SAS involves creating binary variables to represent categories or groups in a dataset. This is useful for statistical analysis or predictive modeling.

An example of creating dummy variables in SAS is as follows: Suppose we have a dataset with a variable "Gender" that has two categories - "Male" and "Female". To create dummy variables for this variable, we can use the SAS "Proc Format" procedure to assign numeric values (0 and 1) to each category and then use the "Proc Transpose" procedure to convert the variable into two binary variables - "Male" and "Female" with values of 0 and 1 indicating the presence or absence of each category. This allows us to include gender as a predictor variable in our analysis.

Create Dummy Variables in SAS (With Example)

A is a type of variable that we create in regression analysis so that we can represent a categorical variable as a numerical variable that takes on one of two values: zero or one.

For example, suppose we have the following dataset and we would like to use *age* and *marital status* to predict *income*:

Income	Age	Marital Status
$45,000	23	Single
$48,000	25	Single
$54,000	24	Single
$57,000	29	Single
$65,000	38	Married
$69,000	36	Single
$78,000	40	Married
$83,000	59	Divorced
$98,000	56	Divorced
$104,000	64	Married
$107,000	53	Married

To use *marital status* as a predictor variable in a regression model, we must convert it into a dummy variable.

Since it is currently a categorical variable that can take on three different values ("Single", "Married", or "Divorced"), we need to create *k*-1 = 3-1 = 2 dummy variables.

To create this dummy variable, we can let "Single" be our baseline value since it occurs most often. Thus, here's how we would convert *marital status* into dummy variables:

Income	Age	Marital Status
$45,000	23	Single
$48,000	25	Single
$54,000	24	Single
$57,000	29	Single
$65,000	38	Married
$69,000	36	Single
$78,000	40	Married
$83,000	59	Divorced
$98,000	56	Divorced
$104,000	64	Married
$107,000	53	Married

Income	Age	Married	Divorced
$45,000	23	0	0
$48,000	25	0	0
$54,000	24	0	0
$57,000	29	0	0
$65,000	38	1	0
$69,000	36	0	0
$78,000	40	1	0
$83,000	59	0	1
$98,000	56	0	1
$104,000	64	1	0
$107,000	53	1	0

The following example shows how to create dummy variables for this exact dataset in SAS.

Example: Creating Dummy Variables in SAS

First, let's create the following dataset in SAS:

```
/*create dataset*/
data original_data;
input income age status $;
datalines;
45 23 single
48 25 single
54 24 single
57 29 single
```

```
65 38 married
69 36 single
78 40 married
83 59 divorced
98 56 divorced
104 64 married
107 53 married
;
run;

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

Obs	income	age	status
1	45	23	single
2	48	25	single
3	54	24	single
4	57	29	single
5	65	38	married
6	69	36	single
7	78	40	married
8	83	59	divorced
9	98	56	divorced
10	104	64	married
11	107	53	married

Next, we can use two IF-THEN-ELSE statements to

create dummy variables for the status variable:

```
/*create new dataset with dummy variables*/
data new_data;
set original_data;
if status = "married" then married = 1;
else married = 0;
if status = "divorced" then divorced = 1;
else divorced = 0;
run;

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

Obs	income	age	status	married	divorced
1	45	23	single	0	0
2	48	25	single	0	0
3	54	24	single	0	0
4	57	29	single	0	0
5	65	38	married	1	0
6	69	36	single	0	0
7	78	40	married	1	0
8	83	59	divorced	0	1
9	98	56	divorced	0	1
10	104	64	married	1	0
11	107	53	married	1	0

Notice that the values for the two dummy variables

(married and divorced) match the values we calculated in the introductory example.

We could then use these dummy variables in a if we'd like since they're both numeric.

Additional Resources

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