How do I interpret the parameter estimates for dummy variables in regression or glm?

Authored by

stats writer

June 30, 2024

The interpretation of parameter estimates for dummy variables in regression or glm refers to the process of understanding the relationship between a categorical variable and the outcome variable in a statistical model. Dummy variables are used to represent categories in a regression or glm model, and the parameter estimates associated with these variables indicate the effect of each category on the outcome variable. By interpreting these estimates, one can determine the strength and direction of the relationship between the categorical variable and the outcome variable. This process is important in understanding the significance of different categories and their impact on the overall model.

How do I interpret the parameter estimates for dummy variables in regression or glm? | SPSS FAQ

Consider this simple data file that has nine subjects (sub) in three groups (iv) with a score on the outcome or dependent variable (dv).

data list list / sub iv dv.
begin data
1 1 48
2 1 49
3 1 50
4 2 17
5 2 20
6 2 23
7 3 28
8 3 30

9 3 32

end data.

Below we use the means command to find the overall mean and the means for the three groups.

means tables = dv by iv.

As we see below, the overall mean is 33, and the means for groups 1, 2 and 3 are 49, 20 and 30 respectively.

	Cases					
Included	**Excluded**	**Total**				
N	**Percent**	**N**	**Percent**	**N**	**Percent**	
DV * IV	9	100.0%	0	.0%	9	100.0%

IV	Mean	N	Std. Deviation
1.00	49.0000	3	1.00000
2.00	20.0000	3	3.00000
3.00	30.0000	3	2.00000
Total	33.0000	9	12.89380

Let's run a standard ANOVA on these data using glm.

glm dv by iv.

The results of the ANOVA are shown below.

	N	
IV	**1.00**	3
2.00	3	
3.00	3	

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1302.000(a)	2	651.000	139.500	.000
Intercept	9801.000	1	9801.000	2100.214	.000
IV	1302.000	2	651.000	139.500	.000
Error	28.000	6	4.667		
Total	11131.000	9			
Corrected Total	1330.000	8			
a R Squared = .979 (Adjusted R Squared = .972)					

Now, let's take this information we have found and relate it to the results that we get when we run a similar analysis using dummy coding. Let's make a data file called dummy2 that has dummy variables called iv1 (1 if iv=1), iv2 (1 if iv=2) and iv3 (1 if iv=3). Note that iv3 is not really necessary, but it could be useful for further exploring the meaning of dummy variables. We will then use the regression command to predict dv from iv1 and iv2.

```
compute iv1 = 0.
if iv = 1 iv1 = 1.
compute iv2 = 0.
if iv = 2 iv2 = 1.
compute iv3 = 0.
if iv = 3 iv3 = 1.
execute.

regression
/dependent = dv
/method = enter iv1 iv2.
```

The output is shown below.

Model	Variables Entered	Variables Removed	Method
1	IV2, IV1(a)	.	Enter
a All requested variables entered.			
b Dependent Variable: DV			

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.989(a)	.979	.972	2.16025
a Predictors: (Constant), IV2, IV1				

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	**Regression**	1302.000	2	651.000	139.500	.000(a)
Residual	28.000	6	4.667			

Total	1330.000	8				
a Predictors: (Constant), IV2, IV1						
b Dependent Variable: DV						

	Unstandardized Coefficients	Standardized Coefficients	t	Sig.		
Model	**B**	**Std. Error**	**Beta**			
1	**(Constant)**	30.000	1.247		24.054	.000
IV1	19.000	1.764	.737	10.772	.000	
IV2	-10.000	1.764	-.388	-5.669	.001	
a Dependent Variable: DV						

First, note that from the ANOVA using the glm command
that the F-value was 139.5 and for the regression using the regression
command the F-value (for the model) is also 139.5. This illustrates that the overall test of the model
using regression is really the same as doing an ANOVA.

After the
ANOVA table, there is a table entitled Coefficients. What is the interpretation of the values listed there, the 30, 19 and
-10? Notice how we have iv1 and iv2 that refer to group 1 and group 2, but we did not include any dummy

variable referring to group 3.
Group 3 is
often called the omitted group or reference group.
Recall that the means of the 3 groups were 49, 20 and 30 respectively. The
intercept
term is the mean of the dependent variable, which we called dv, for the
omitted group, and indeed the parameter estimate
(in the column B) from the output is the mean of group 3, 30. The parameter estimate for
iv1
is the mean of the dependent variable, dv, for group 1 minus the mean of
the dependent variable for group 3, 49 - 30 = 19, and indeed that is the
parameter estimate for iv1. Likewise, the parameter estimate for
iv2
is the mean of the dependent variable for group 2 minus the mean of the
dependent variable for group 3, 20 - 30 = -10, the parameter estimate for
iv2.

So, in summary:

Intercept	mean of group 3 (mean of omitted group)
iv1	mean of group 1 - group 3 (omitted group)
iv2	mean of group 2 - group 3 (omitted group)

Try running this example, but use iv2 and iv3 using regression (making group 1 the omitted group) and see what happens.

Finally, consider how the parameter estimates can be used in the regression model to obtain the means for the groups (the predicted values).

The regression model is

Ypredicted = 30 + iv1*19 + iv2*-10

For group 1: Ypredicted = 30 + 1 * 19 + 0 * -10 = 49
For group 2: Ypredicted = 30 + 0 * 19 + 1 * -10 = 20
For group 3: Ypredicted = 30 + 0 * 19 + 0 * -10 = 30

As you see, the regression formula predicts that each group will have the mean value of its group.

You can also perform the same analysis using

glm. The print = parameter subcommand tells SPSS to print the regression coefficients.

glm dv with iv1 iv2 /print = parameter.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1302.000(a)	2	651.000	139.500	.000
Intercept	2700.000	1	2700.000	578.571	.000
IV1	541.500	1	541.500	116.036	.000
IV2	150.000	1	150.000	32.143	.001
Error	28.000	6	4.667		
Total	11131.000	9			
Corrected Total	1330.000	8			
a R Squared = .979 (Adjusted R Squared = .972)					

	B	Std. Error	t	Sig.	95% Confidence Interval	
Parameter	**Lower Bound**	**Upper Bound**				
Intercept	30.000	1.247	24.054	.000	26.948	33.052
IV1	19.000	1.764	10.772	.000	14.684	23.316
IV2	-10.000	1.764	-5.669	.001	-14.316	-5.684