Q-STEP SPSS 'HOW TO' GUIDES: Recoding categorical variables into dummy variables in spss

Creator: Joshua Townsley

Say we wanted to run a regression model in SPSS with a categorical independent variable:

Normally, we would interpret each IV's coefficient in a regression as the change in the DV for a one unit increase in that particular IV. But, if we have a categorical variable such as Ethnicity or Age Group, then what would a one unit increase in that variable actually mean? It's either nonsensical or tricky to interpret.

So instead, we have to break that categorical variable up into 'dummy variables' for each particular value/answer. This is a nuisance, but it's really important for any running regression in SPSS that uses a categorical Independent Variable.

E.g. Let's say we have a variable – Age Group – coded as follows:

1 = Under 18 2 = 18-25 3 = 26-35 4 = 36-45 5 = 46-55 6 = 56-65 7 = 66+

We'll have to break it up into 'Dummy Variables' (coded 1 for 'yes'/0 for 'no') for each particular age group. E.g. AgeGroup1 - Under 18 (yes or no), AgeGroup2 - 18-25 (yes or no), AgeGroup3

- 26-35 (yes or no), and so on. So, we will end up with as many new variables as there are categories in our categorical variables (in this case, 7).

There are basically two ways to do this – an easier way and a trickier way. This guide runs through each one below, but the easy one is of course better (but may not be available for older versions of SPSS).









1. Method 1

Go through the following menus on SPSS:

Transform > Create Dummy Variables

1. Put your categorical independent variable into the top right box, and enter an intuitive name for the dummy variables into the 'Root Names' box

/ariables:	Create Dummy Variables for:
🖋 id 🛛 👔	🚓 ageGroup
🖉 wt_full_W10	
wave10	*
ស country	
& generalElectionVote	
a generalElectionVoteOth	Main Effect Dummy Variables
generalElectionCertainty	✓ Create main-effect dummies
o partyldstrength	Root Names (One Per Selected Variable):
Dummy Variable Labels	agegroup
Use value labels	Macro Name:
Use values	
Value Order	Two-Way Interactions
Ascending	Create dummies for all two-way interactions
Opescending	Root Name:
Macros	Macro name:
Omit first dummy category from macro definitions	
Note: It is conventional to start macro names with !.	Three-Way Interactions
Massurement Level Usage	Create dummies for all three-way interactions
Do not create dummies for scale variable values	Root Name:
Create dummies for all variables	
Create dummes for an variables	Macro name:
This dialog requires the Python Essentials	

2. Select 'OK' to create the dummy variables. They will then appear at the bottom of the variable list:



Nam Nam Nam Nam Nam Nam Nam Nam	me nSch ondS rivSc hne /elsh	Type Numeric Numeric Numeric Numeric Numeric Numeric	Width 10 10 10 10 21 17	Decimals 0 0 0 0 0 0 0 0 0 0	Label Went to privat Went to privat Never went to Private school: Speaks Welsh? Education level	Values (0, No) (0, No) (0, No) (0, No) (0, No) (0, No cuali	Missing None None None None	Columns 8 8 8 8 8 8 8 8	Align Right Right Right Right Right Right	Measure Nominal Nominal Nominal Nominal Nominal	Role Input Input Input Input Input
386 privPrim 387 privSeco 388 neverPri 389 RPrivSch 390 speakWe 391 edlevel 392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou	nSch ondS rivSc hne (elsh	Numeric Numeric Numeric Numeric Numeric Numeric	10 10 10 21 17	0 0 0 0 0 0	Went to privat Went to privat Never went to Private school: Speaks Welsh? Education level	(0, No) (0, No) (0, No) (0, No) (0, No)	None None None None	8 8 8 8 8	理 Right 理 Right 理 Right 理 Right 理 Right	 Nominal Nominal Nominal Nominal Nominal Nominal 	Input Input Input Input Input
387 privšeco 388 neverPri 389 RPrivšch 390 speakWe 391 edlevel 392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou	ondS rivSc hne /elsh	Numeric Numeric Numeric Numeric Numeric	10 10 10 21 17	0 0 0 0 0	Went to privat Never went to Private school: Speaks Welsh? Education level	(0, No) (0, No) (0, No) (0, No)	None None None	8 8 8 8	Right Right Right Right Right	 Nominal Nominal Nominal Nominal 	Input Input Input
388 neverPri 389 RPrivSch 390 speakWe 391 edlevel 392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou	rivSc hne /elsh	Numeric Numeric Numeric Numeric Numeric	10 10 21 17	0 0 0 0	Never went to Private school: Speaks Welsh? Education level	(0, No) (0, No) (0, No)	None None None	8 8 8	Right Right Right	& Nominal & Nominal & Nominal	> Input > Input > Input
389 RPrivSch 390 speakWe 391 edlevel 392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou	hne /elsh	Numeric Numeric Numeric Numeric	10 21 17	0 0 0	Private school: Speaks Welsh? Education level	(0, No) (0, No)	None	8	Right	💫 Nominal	> Input
390 speakWe 391 edlevel 392 lr_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou	/elsh	Numeric Numeric Numeric	21	0	Speaks Welsh? Education level	(0, No) (0, No guali	None	8	Right	💫 Nominal	S Input
391 edlevel 392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou		Numeric Numeric	17	0	Education level	0. No quali					
392 Ir_scale 393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou		Numeric	â			tar' un darque	None	8	Right Right	💑 Nominal	S Input
393 al_scale 394 Degree 395 agegrou 396 agegrou 397 agegrou			9	2	Left-right scale	{.00, Left}	None	8	Right	scale 🧳	🔪 Input
394 Degree 395 agegrou 396 agegrou 397 agegrou	e	Numeric	13	2	Authoritarian/1	{.00, Liberta	None	8	Right	scale 🤣	🔪 Input
395 agegrou 396 agegrou 397 agegrou		Numeric	8	2	Degree?	None	None	10	Right	🚓 Nominal	S Input
396 agegrou 397 agegrou	up_1	Numeric	8	2	ageGroup=Und	None	None	12	Right	💦 Nominal	S Input
397 agegrou	up_2	Numeric	8	2	ageGroup=18	None	None	12	Right	🕹 Nominal	🔪 Input
	up_3	Numeric	8	2	ageGroup=26	None	None	12	Right	💦 Nominal	🔪 Input
398 agegrou	up_4	Numeric	8	2	ageGroup=36	None	None	12	Right	💫 Nominal	🔪 Input
399 agegrou	up_S	Numeric	8	2	ageGroup=46	None	None	12	Right	💦 Nominal	🔪 Input
400 agegrou	up_6	Numeric	8	2	ageGroup=56	None	None	12	Right	Nominal	🔪 Input
401 agegrou		Numeric	8	2	ageGroup=66+	None	None	12	Right	🛞 Nominal	> Input

3. For each new dummy variable, click on the Values box and create the value labels as appropriate for its values. E.g.:

	Value Labels		{0, No}	None
t - Value Labels			{0, No}	Non
o Value			{0, No}	Non
value:		Spelling	{0, No}	Non
Label:			{0, No}	Non
e	.00 = "Else"		{0, No quali	Non
al Add	1.00 = "Under 18"		{.00, Left}	Non
Change			{.00, Liberta	. Non
			None	None
Remove			None	None
Help		Cancel OK		

(And so on for each dummy variable)



4. Once all of the dummy variables have been assigned Value Labels, you can then enter these into the Regression.

Note: when you do this, you enter all of the dummy variables EXCEPT ONE. This is important, because the one you leave out is the category that you compare the other dummy variables' coefficients to. E.g. if you enter the dummy variables for all the age group categories except the Under 18s, then the coefficients that they produce are telling you the difference between that age group compared to Under 18s.

(This YouTube video walks through this interpretation: https://www.youtube.com/watch? v=AJ88dheUieY).

2. Method 2

Go through the following menus on SPSS:

Transform > Recode into different variables

1. Put Age Group into the main box, then select 'Old and New Values'. To create the first dummy variable (for Under 18s), enter 1 into the value box on the left hand side, and 1 into the New Value box on the right hand side. Then, select 'All other values' on the left side, and enter '0' on the right side. It will end up looking like this:

Old Value	New Value
🔿 Value:	Value:
	System-missing
System-missing	Copy old value(s)
System- or user-missing	Old> New:
Kange:	1> 1 ELSE> 0
through	Add
	Change
Range, LOWEST through value:	Remove
Range, value through HIGHEST:	
	Output variables are strings Width: 8
All other values	Convert numeric strings to numbers ('5'->5)
2	Cancel



1. Select 'Continue', then on the window, enter a name and label for thenew dummy variable. Call it something intuitive, e.g. 'Under18' (to account for the fact that we have coded it so Under 18s = 1, and all other groups= 0).

profile_past_vot profile_turnout profile_eurefvote profile_eureftur euRefLA pano onscode gender marital	ageGroup> ?	Output Variable Name: Under18 Label: Under 18? Change
b housing gor education pcon profile educatio	Old and New Values	lition)

3. Find the new dummy variable at the bottom of the main variable list window, and create the value labels as appropriate for its values. E.g.:

Val	ue:		Spelling	{0, No} {0, No}	Non
e	Add	.00 = "Else" 1.00 = "Under 18"		{0, No} {0, No quali {.00, Left}	None
/	Change			{.00, Liberta None	Non
	Remove			None	Non



4. Repeat the above for every value of the original Age Group variable, but coded respectively for each category. E.g. 18-25 = 1, Else = 0; then 26-35 = 1, Else = 0; and so on.

E.g. for 18-25:

Recode so that 18-25s are '1', and all other values are '0':

Old Value	New Value
O Value:	• Value:
	System-missing
System-missing	Copy old value(s)
System- or user-missing	Old> New:
Range:	2> 1 ELSE> 0
through	Add
Range, LOWEST through value:	Change Remove
Range, value through HIGHEST:	Output variables are strings Width: 8
All other values	Convert numeric strings to numbers ('5'->5)
	Constant Constant

Set the name and label to something intuitive:

<pre> id wt_full_W10 wave10 country turnoutUKGeneral comecalElection) </pre>	Numeric Variable -> Output Variable: ageGroup> AgeGroup18to25	Output Variable Name: AgeGroup18to25 Label: 18 to 25?
generalElectionV generalElectionV generalElectionC partyldStrength partyld partyldOth partyldSqueeze partyldSqueeze countryOfBirth	Old and New Values If (optional case selection cond	Change dition)



Go to the new AgeGroup18to25 variable and change the corresponding Value Labels:

Value:		Spelling
_abel:		
	.00 = "Else"	
Add	1.00 = "18-25"	
Change		
Remove		

...and repeat for the remaining categories.

5. Once all of the categories have a respective dummy variable, you can then enter these into the Regression. Note: when you do this, you enter all of the variables EXCEPT ONE. This is important, because the one you leave out is the category that you compare the other dummy variables' coefficients to. E.g. if you enter the dummy variables for all the age group categories except the Under 18s, then the coefficients that they produce are telling you the difference between that age group compared to Under 18s.

(This YouTube video walks through this interpretation: https://www.youtube.com/watch? v=AJ88dheUieY in a *relatively* simple way).

