

Dummy Variable Coding

Points taken from notes 8d Regression with One Qualitative Predictor

1. Coding of a Dummy Variable, also known as an Indicator Variable

For binary variables scores = 1 or 0

Examples

- Sex: Female = 1, Male = 0
- Pass/Fail Status: Pass = 1, Fail = 0
- Drug Usage: Use = 1, non-use = 0

For multcategory variables, $c - 1$ = number of dummy variables needed in the regression equation, where c = number of categories in variable

Examples:

- Sex: Female and Male, $c = 2$, so $2 - 1 = 1$ dummy variable required in regression
- Race/Ethnicity: Asian, Black, Hispanic, White, $c = 4$ so $4 - 1 = 3$ dummy variables required for regression
- Three teaching strategies compared: Strategy A, B, and C, so $c = 3$, thus 2 dummy variables...
- If a categorical variable had 7 categories, then $c = 7 - 1 = 6$ dummy variables...

2. Missing Data

How is missing data identified in your data file?

- blank cell, e.g., 
- period, e.g., 
- numeric identifier such as -9 or -99, or 99, e.g., 

With SPSS and other statistical packages, this is easily handled (as shown below), but with Excel, it can be challenging.

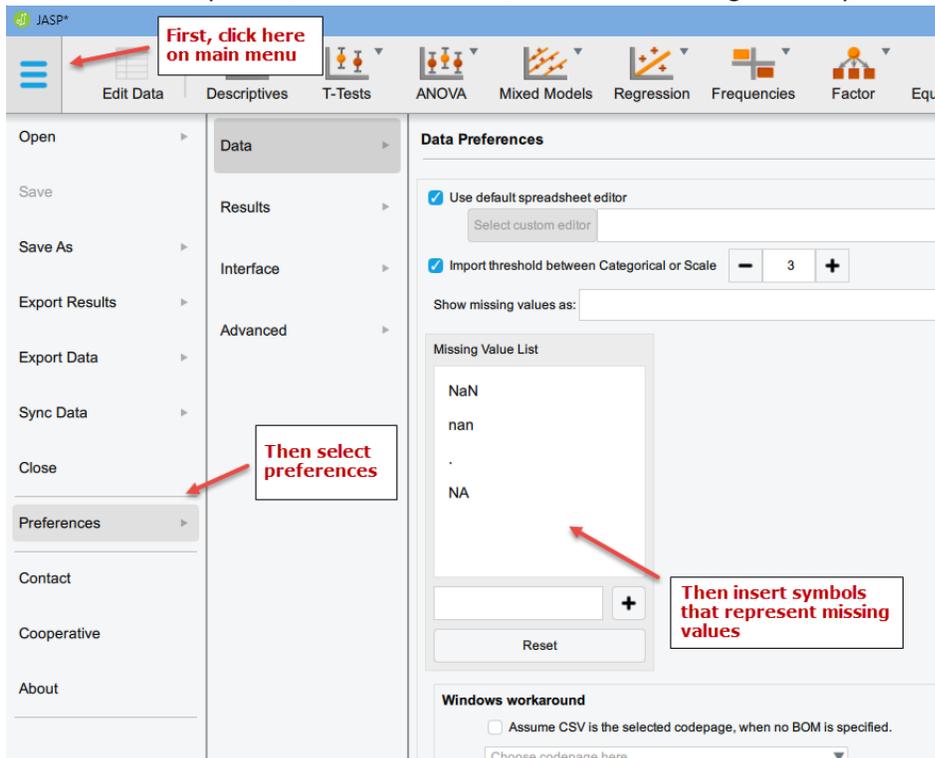
3. Creating Dummy Variables

SPSS

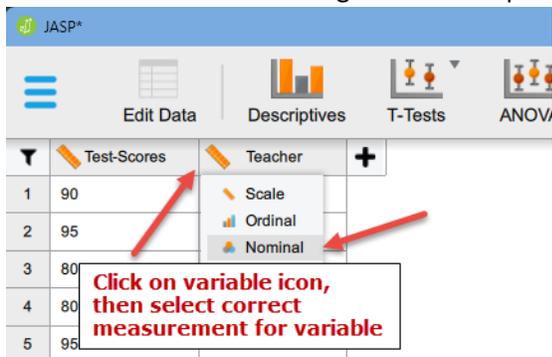
- Create dummy variables, use
 - Transform -> Recode (into different variable)
 - assign Output Variable a name, select Change
 - assign Old and New values
 - Old value (e.g. 3) = 1
 - System- or User-missing = System-missing
 - 99 = System-missing
 - All other values = 0
 - select Continue
 - select OK
 - Repeat the above process for each category of the variable until all needed dummy variables are created for the categorical variable of interest.
- Check frequencies of old variable and dummy variables to ensure correct count of values transferred and that correct missing data still remain

JASP

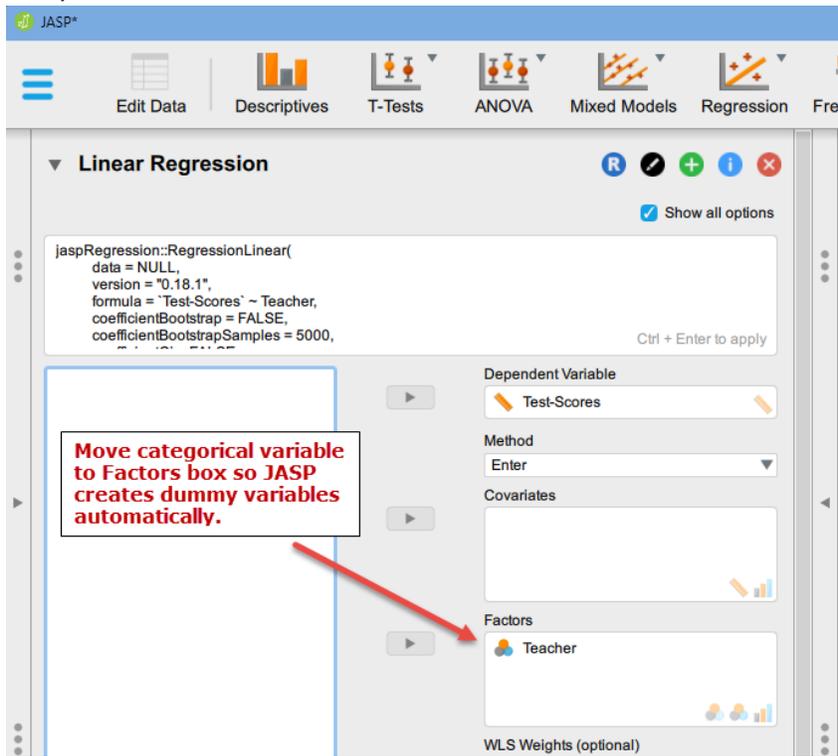
First, select Main Menu symbols, then Preferences, then enter missing values symbols



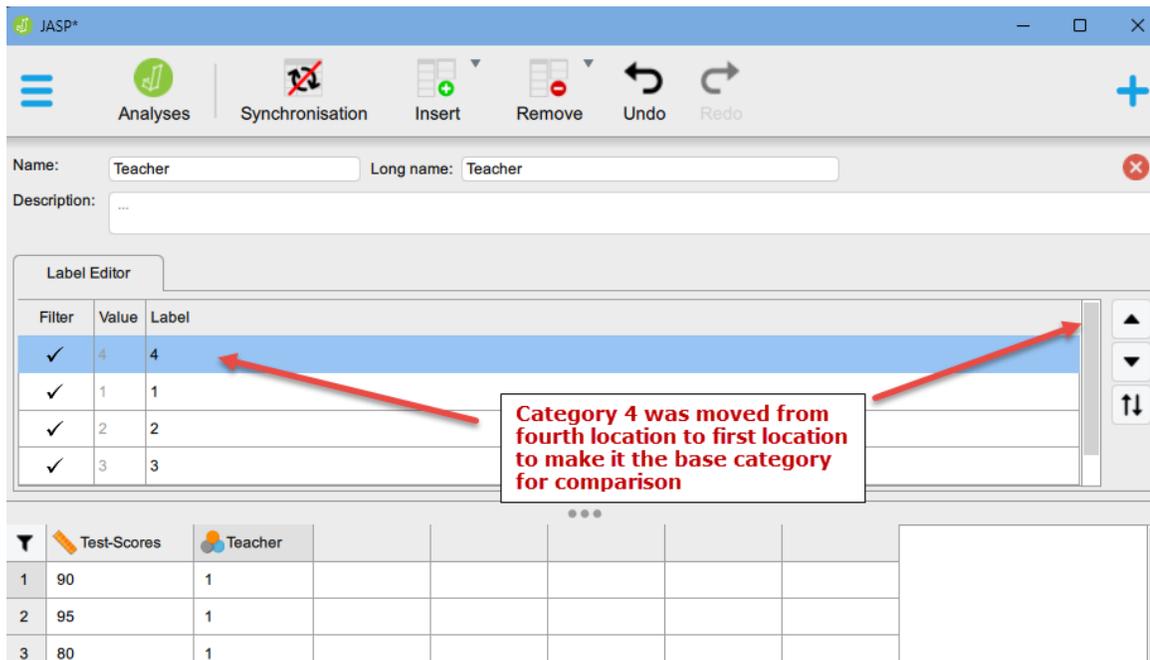
Next ensure the variable that is categorical and requires dummy variables is identified by JASP as nominal or ordinal



Then start Regression analysis and move the categorical variable into the Factors box so JASP creates dummy variables automatically.



How to change the base category for comparison in JASP? By default JASP treats the first category as the comparison group. This can be changed by reordering the categories for the variable in JASP data editor.



Excel or Google Sheets

See example spreadsheet and formulas to convert variable missing value symbols to other identifiable symbols. The formulas presented work in both Excel and Google Sheets.

	A	B	C	D	E
1	test_scores	teacher	=IF(B2=1,1,0)	=IF(ISBLANK(B2),99,IF(B2=1,1,0))	=IF(ISNUMBER(VALUE(B2))=TRUE,B2,99)
2	90	1	1	1	1
3	95	1	1	1	1
4	80	1	1	1	1
5	80	2	0	0	2
6	95	2	0	0	2
7	75	2	0	0	2
8	85	3	0	0	3
9	75	3	0	0	3
10	90	3	0	0	3
11	75	4	0	0	4
12	85	4	0	0	4
13	70	4	0	0	4
14	65	99	0	0	99
15	85	.	0	0	99
16	95		0	99	0
17				Use if missing is blank	Use if missing is ". " to change missing market to a specific number
18					
19					
20					