Neuendorf

Dummy Coding and Effect(s) Coding

For a nominal variable with c categories, you may create up to c-1 dummy or effect variables.
For all equations below, assume the all-standardized situation.

Imagine a four-group religion variable: All respondents are either self-declared Christian, Muslim, Jewish, or Other. From this single nominal variable, three dummies (or effect coded variables) may be created.

DUMMY CODING

	D_1	D_2	D_3	
Christian	1	0	0	
Muslim	0	1	0	
Jewish	0	0	1	
Other	0	0	0	– [The chosen "reference category"]

Christian
$$\begin{aligned} Y' &= \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \\ &= \beta_1 D_1 + 0 + 0 \\ &= \beta_1 D_1 \\ &= \beta_1 && \text{[because } D_1 = 1\text{]} \end{aligned}$$

Muslim
$$\begin{aligned} Y' &= \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \\ &= 0 + \beta_2 D_2 + 0 \\ &= \beta_2 D_2 \\ &= \beta_2 && \text{[because } D_2 = 1\text{]} \end{aligned}$$

Jewish
$$\begin{aligned} Y' &= \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \\ &= 0 + 0 + \beta_3 D_3 \\ &= \beta_3 D_3 \\ &= \beta_3 && \text{[because } D_3 = 1\text{]} \end{aligned}$$

Other
$$\begin{aligned} Y' &= \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

Original Category	Expected value (Y')
Christian	β_1
Muslim	β_2
Jewish	β_3
Other	0

Hence:
- β_1 is the difference between the expected values of Christian and Other,
- β_2 is the difference between the expected values of Muslim and Other, and
- β_3 is the difference between the expected values of Jewish and Other.

Compare the meaning of the <u>partial</u> regression coefficients with the simple r's between Y and D1, D2, and D3.

EFFECT CODING

	E_1	E_2	E_3	
Christian	1	0	0	
Muslim	0	1	0	
Jewish	0	0	1	
Other	-1	-1	-1	– [The chosen "reference category"]

Christian Y' =
Muslim Y' =
Jewish Y' =

All three same as for dummy coding

Other $Y' = \beta_1 E_1 + \beta_2 E_2 + \beta_3 E_3$
$= \beta_1(-1) + \beta_2(-1) + \beta_3(-1)$
$= 0 - \beta_1 - \beta_2 - \beta_3$

Original Category	Expected value (Y')
Christian	β_1
Muslim	β_2
Jewish	β_3
Other	$0 - \beta_1 - \beta_2 - \beta_3$

Hence:

β_1 is ½ the difference between the expected values of Christian and the other three groups (Muslim, Jewish, Other); the sum of the coefficients for those three groups is $-\beta_1$

β_2 is ½ the difference between the expected values of Muslim and the other three groups (Christian, Jewish, Other); the sum of the coefficients for those three groups is $-\beta_2$

β_3 is ½ the difference between the expected values of Jewish and the other three groups (Christian, Muslim, Other); the sum of the coefficients for those three groups is $-\beta_3$

Again, compare the meaning of the <u>partial</u> regression coefficients with the simple r's between Y and E1, E2, and E3.

IN SUM, THEN:

1. For dummy coding, each test of a β indicates the difference between the dummy group and the "reference category" (the category/group that got all 0s).

2. For effect coding, each test of a β indicates the difference between the effect group and all other groups combined.

3. Also, to test for the impact of a set of dummies or effect variables, include all as a block and look at the size and significance of the R^2 change.

3/19