Neuendorf

**Dummy Coding and Effects Coding**

For a nominal variable with c categories, you may create up to c-1 dummy or effects 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 Buddhist. From this single nominal variable, three dummies (or effects coded variables) may be created.

**DUMMY CODING**

<table>
<thead>
<tr>
<th>Dummy</th>
<th>Christian</th>
<th>Muslim</th>
<th>Jewish</th>
<th>Buddhist</th>
</tr>
</thead>
<tbody>
<tr>
<td>D₁</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D₂</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D₃</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

For **Christian**:

\[ Y' = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \]

\[ = \beta_1 + 0 + 0 \]

\[ = \beta_1 \]

[because D₁ = 1]

For **Muslim**:

\[ Y' = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \]

\[ = 0 + \beta_2 + 0 \]

\[ = \beta_2 \]

[because D₂ = 1]

For **Jewish**:

\[ Y' = \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 \]

\[ = 0 + 0 + \beta_3 \]

\[ = \beta_3 \]

[because D₃ = 1]

**Hence:**

\[ \beta_1 \] is the difference between the expected values of Christian and Buddhist,

\[ \beta_2 \] is the difference between the expected values of Muslim and Buddhist, and

\[ \beta_3 \] is the difference between the expected values of Jewish and Buddhist.

Compare the meaning of the partial regression coefficients with the simple r’s between Y and D₁, D₂, and D₃.
**EFFECTS CODING**

<table>
<thead>
<tr>
<th></th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christian</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Muslim</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Jewish</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Buddhist</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

[The chosen “reference category”]

- Christian $Y' = \beta_1 E_1$
- Muslim $Y' = \beta_2 E_2$
- Jewish $Y' = \beta_3 E_3$

- **Buddhist**
  
  $Y' = \beta_1(-1) + \beta_2(-1) + \beta_3(-1)$
  
  $= 0 - \beta_1 - \beta_2 - \beta_3$

**Original Category**

<table>
<thead>
<tr>
<th>Category</th>
<th>Expected value ($Y'$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christian</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>Muslim</td>
<td>$\beta_2$</td>
</tr>
<tr>
<td>Jewish</td>
<td>$\beta_3$</td>
</tr>
<tr>
<td>Buddhist</td>
<td>0 - $\beta_1$ - $\beta_2$ - $\beta_3$</td>
</tr>
</tbody>
</table>

Hence:

- $\beta_1$ is $\frac{1}{2}$ the difference between the expected values of Christian and the other three groups (Muslim, Jewish, Buddhist); the sum of the coefficients for those three groups is $-\beta_1$
- $\beta_2$ is $\frac{1}{2}$ the difference between the expected values of Muslim and the other three groups (Christian, Jewish, Buddhist); the sum of the coefficients for those three groups is $-\beta_2$
- $\beta_3$ is $\frac{1}{2}$ the difference between the expected values of Jewish and the other three groups (Christian, Muslim, Buddhist); the sum of the coefficients for those three groups is $-\beta_3$

Again, compare the meaning of the partial 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 $\beta$ indicates the difference between the dummy group and the “reference category.”
2. For effects coding, each test of a $\beta$ indicates the difference between the effects group and all other groups pooled.
3. Also, to test for the impact of a set of dummies or effects variables, include all as a block and look at the size and significance of the R$^2$ change.

2/09