Three Types of Coding for Categorical Variables

Single Variable with Three Categories

(with equal *n*'s per category/unit vector included)

		Dummy			Effect			Orthogonal		
Case	Group	U_1	D_2	D_3	U_1	D_2	D_3	U_1	D_2	D_3
1	1	1	0	0	1	-1	-1	1	-1	-½
2	1	1	0	0	1	-1	-1	1	-1	-½
3	1	1	0	0	1	-1	-1	1	-1	-½
4	2	1	1	0	1	1	0	1	1	-½
5	2	1	1	0	1	1	0	1	1	-½
6	2	1	1	0	1	1	0	1	1	-½
7	3	1	0	1	1	0	1	1	0	1
8	3	1	0	1	1	0	1	1	0	1
9	3	1	0	1	1	0	1	1	0	1

Hypotheses:

For dummy coding:

b_2 - Does group 2 differ from group 1?
b_3 - Does group 3 differ from group 1?

For effect coding:

b_2 - Does group 2 differ from the mean of the group means?
b_3 - Does group 3 differ from the mean of the group means?

For orthogonal coding:

b_2 - Does group 2 differ from group 1?
b_3 - Does group 3 differ from the mean of groups 1 and 2?

Tests: If the partial regression coefficient corresponding to a given variable equals zero, there is no difference.

N.B.:

1. For effect coding the effects must sum to zero: hence the difference between group 1 and the mean of group means is equal to minus the sum of the estimated effects (i.e., zero minus the sum of the two regression coefficients.
2. For orthogonal contrasts: (a) the weights within any contrast (vector) must sum to zero and (b) the weights for any pair of contrasts must be orthogonal (the dot product of the two vectors must be zero).