PLE LINEAR REGRESSION VIEWPOINTS
LUME 15, NUMBER 2, WINTER 1987

The Use of Nonsence Coding with ANOVA Situations

John D. Williams
The University of North Dakota

Summary: Nonsense coding systems can be constructed that retain outcomes regarding R^2 values, F values and multiple comparison tests. Nonsense coding highlights the flexibility of coding ANOVA problems to be analyzed by multiple linear regression procedures; however, no additional analytic power appears to be gained from their use.

Characteristic Coding Compared to Nonsense Coding

Most coding systems for accomplishing ANOVA solutions by multiple linear regression use some variant of characteristic coding (binary coding/dummy coding) with the use of 1's or 0's, depending upon group membership, or contrast coding, which uses 1's, 0's and -1's (see Williams, 1974a). The use of orthogonal contrasts deviates from this usage, including orthogonal polynomials, but none of these systems allow arbitrariness in their coding process.

On the other hand, Cohen and Cohen (1975) assert that regression solutions can be accomplished through the use of "nonsense" coding, though they neither give directions nor examples of this process. Thus, an example of nonsense coding is provided here. The data are taken from Williams (1974b, p. 43, problem 5.3). See Table 1.

Table 1

Sample Data for ANOVA Problem

Group One	Group Two	Group Three	Group Four	Group Five
19	20	13	12	22
18	19	12	8	20
15	16	10		19
13	16	10		19
8	14	10		15
5	14			
	13			

The data in Table 1 are clearly from unequal sized groups; the intent is to show outcomes that have generality beyond equal cell sized situations. First, to accomplish a characteristic coding of this data:

Y = the criterion score;

X_1 = 1 if a member of Group One, 0 otherwise;

X_2 = 1 if a member of Group Two, 0 otherwise;

X_3 = 1 if a member of Group Three, 0 otherwise;

X_4 = 1 if a member of Group Four, 0 otherwise; and

X_5 = 1 if a member of Group Five, 0 otherwise.

Table 2 shows these values for the data in Table 1.

Table 2

Characteristic Coding (1 or 0) for Data in Table 1

Y	X_1	X_2	X_3	X_4	X_5
19	1	0	0	0	0
18	1	0	0	0	0
15	1	0	0	0	0
13	1	0	0	0	0
8	1	0	0	0	0
5	1	0	0	0	0
20	0	1	0	0	0
19	0	1	0	0	0
16	0	1	0	0	0
16	0	1	0	0	0
14	0	1	0	0	0
14	0	1	0	0	0
13	0	1	0	0	0
13	0	0	1	0	0
12	0	0	1	0	0
10	0	0	1	0	0
10	0	0	1	0	0
10	0	0	1	0	0
12	0	0	0	1	0
8	0	0	0	1	0
22	0	0	0	0	1
20	0	0	0	0	1
19	0	0	0	0	1
19	0	0	0	0	1
15	0	0	0	0	1

Next five different linear models can be defined to complete an analysis by multiple linear regression:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1; \quad (1)$$

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_5X_5 + e_1; \quad (2)$$

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + b_5X_5 + e_1; \quad (3)$$

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + b_5X_5 + e_1; \quad (4)$$

and $$Y = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + e ; \quad (5)$$

Equations 1 thru 5 are reparameterizations of each other and are reparameterizations of

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + e_1. \qquad (6)$$

The use of equations 1 thru 5 require the use of a unit vector for solution (commonly a part of typical multiple use multiple lienar regression programs) and represent solutions that allow psuedo-Dunnett formulations that permit construction of all simple comparisons of means (see Williams, 1976). Also, the b_i's are unique to each equation. Each of the formulations yields $R^2 = .49362$, $F = 4.874$ with df = 4,20 and $p < .05$. A part of the printout is shown in Table 3 for equation 1.

Table 3

Portions of Printout for Multiple Linear Regression for the Sample Data in Table 1 Using 1 or 0 Coding

Variable	Mean	Correlation	Regression Coefficient	Standard Error of Regression Coefficient	Comput t Valu
X_1	.240	-.181	-6.000	2.089	-2.872
X_2	.280	.230	-3.000	2.020	-1.485
X_3	.200	-.392	-8.000	2.182	-3.667
X_4	.080	-.299	-9.000	2.886	-3.118
Criterion	14.400				
Intercept	19.000				

In Table 3, means refer to the proportion in a group for characteristic (1 or 0) coded data. The regression coefficient is the difference between the mean of the particular coded group and the "left-out" group (Group Five). If the regression coefficient is divided by its own standard error, the computed t value is found which can be compared to a table for an appropriate multiple comparison method (e.g., Tukey's test). The correlations in Table 3 represent point-biserial correlations of the group membership variables with the criterion. The criterion is the overall mean for the Y scores, and the intercept (b_0) is the mean of the "left-out" group (Group Five). A reformulation of equation 1 makes these relationships more obvious:

$$Y = \bar{Y}_5 + (\bar{Y}_1-\bar{Y}_5)X_1 + (\bar{Y}_2-\bar{Y}_5)X_2 + (\bar{Y}_3-\bar{Y}_5)X_3 + (\bar{Y}_4-\bar{Y}_5)X_4 + e_1. \quad (7)$$

The set of all simple multiple comparisons, omitting signs and lower diagonal entries is shown in Table 4.

Table 4

Means and Computed t Values for all Simple Comparisons Using Characteristic (1 or 0) Coding

Group	One	Two	Three	Four	Five
Mean	13.00	16.00	11.00	10.00	19.00
One		1.563	.957	1.065	2.872
Two			2.475	2.169	1.485
Three				.348	3.667
Four					3.118

Using Tukey's Test ($p < .05$) a t value of 2.992 is required for significance.

Using Nonsense Coding

Nonsense coding consistent with the characteristic coding process can be accomplished in the following manner:

Let Y = the criterion score

X_1 = a if a member of Group One, b otherwise (a ≠ b);

X_2 = c if a member of Group Two, d otherwise (c ≠ d);

X_3 = e if a member of Group Three, f otherwise (e ≠ f);

X_4 = g if a member of Group Four, h otherwise (g ≠ h); and

X_5 = i if a member of Group Five, j otherwise (i ≠ j).

It can be noted that the solution given earlier is the special case using this notation where a = c = e = g = i = 1 and b = d = f = h = j = 0. As an example of choosing values for a thru j, let a = 7, b = 3, c = 2, d = 9, e = 4, f = 1, g = 8, h = 5, i = 6, and j = 2. Using these values, similar equations were constructed to equations 1 thru 5 and multiple regressions were completed. For the data set itself, see Table 5.

Table 5

Characteristic Coding Using a Nonsense Coding Process for Data in Table 1

Y	X_1	X_2	X_3	X_4	X_5
19	7	9	1	5	2
18	7	9	1	5	2
15	7	9	1	5	2
13	7	9	1	5	2
8	7	9	1	5	2
5	7	9	1	5	2
20	3	2	1	5	2
19	3	2	1	5	2
16	3	2	1	5	2
18	3	2	1	5	2
14	3	2	1	5	2
14	3	2	1	5	2
13	3	2	1	5	2
13	3	9	4	5	2
12	3	9	4	5	2
10	3	9	4	5	2
10	3	9	4	5	2
10	3	9	4	5	2
12	3	9	1	8	2
8	3	9	1	8	2
22	3	9	1	5	6
20	3	9	1	5	6
19	3	9	1	5	6
19	3	9	1	5	6
15	3	9	1	5	6

Using formulations like equations 1 thru 5, each equation yields $R^2 = .49362$, $F = 4.874$, with df = 4,20 and $p < .05$, identically the same as before.

The appearance of other portions of the printout is somewhat changed; a portion of the printout corresponding to equation 1 is shown in Table 6 and can be compared to Table 3.

Table 6

Portions of Printout for Multiple Linear Regression Using Nonsense Coding for the Sample Data in Table 1

Variable	Mean	Correlation	Regression Coefficient	Standard Error of Estimate	Computed t Value
X_1	3.960	-.181	-1.500	.522	-2.872
X_2	7.040	-.230	.429	.289	1.485
X_3	1.600	-.392	-2.667	.727	-3.667
X_4	5.240	-.299	-3.000	.962	-3.118
Criterion	14.400				
Intercept	37.309				

It is by no means obvious what the meaning of the mean, regression coefficient or standard error of estimate are from a cursory glance at the output. However, the correlation coefficients remain point-biserial correlation coefficients of each group membership variable with the criterion even though they are not 1's and 0's. Also, except for sign, the computed t values are identical with those found earlier. Thus, even though much of the output is unfamiliar, important aspects are identical to those found earlier. Actually, the means represent simply the mean values of a variable assigned by our coding scheme; for example, the coding in Group One on X_1 is 3 and .24 of the scores are from this group. The remaining .76 are from other groups and are coded 7. Then .24(3) + .76(7) = 3.96, the mean of X_1. The regression coefficients are part of the least squares process that achieve the same expected values as was found previously, that is, the mean for the group. A rather intractable equation, siimilar to equation 7, relates the means for the nonsense coding situation: $Y = \bar{Y}_5 - \{[b(\bar{Y}_1 - \bar{Y}_5)/ (a - b)] + [d(\bar{Y}_2 - \bar{Y}_5)/ (c - d)] + [f(\bar{Y}_3 - \bar{Y}_5)/(e - f)] [h(\bar{Y}_4 - \bar{Y}_5)/(g - h)]\} + [(\bar{Y}_1 - \bar{Y}_5)/(a - b)]X_1 + [(\bar{Y}_2 - \bar{Y}_5)/(c - d)]X_2 + [\bar{Y}_3 - \bar{Y}_5)/ (e - f)]X_3 + [(\bar{Y}_4 - \bar{Y}_5)/(g - h)]X_4 + e_1$. (8)

The relationship of the regression coefficients to the standard errors of estimate remains proportional so that the computed t values remain identical to those found for the characteristic coding solution.

Contrast Coding with Nonsense Coding

Some researchers prefer to use contrast coding (see Williams, 1974a) to characteristic coding systems, particularly if they are interested in a traditional analysis of variance solution.* A typical contrast coding systems using either a 1 or -1 or 0 is as follows:

*Because the computed t values are directly interpretable as multiple comparisons (see equation 7) characteristic coding solutions would seem to be preferable for testing most hypotheses of interest making the characteristic coding solution not only simpler to achieve but more useful as well.

X_1 = 1 if a member of Group 1, -1 if a member of Group 5, 0 otherwise;

X_2 = 1 if a member of Group 2, -1 if a member of Group 5, 0 otherwise;

X_3 = 1 if a member of Group 3, -1 if a member of Group 5, 0 otherwise; and

X_4 = 1 if a member of Group 4, -1 if a member of Group 5, 0 otherwise.

A nonsense contrast coding can be accomplished as follows:

X_1 = a if a member of Group 1, -a if a member of Group 5, b otherwise ($a \neq b$);

X_2 = c if a member of Group 2, -c if a member of Group 5, d otherwise ($c \neq d$);

X_3 = e if a member of Group 3, -e if a member of Group 5, f otherwise ($e \neq f$); and

X_4 = g if a member of Group 4, -g if a member of Group 5, h otherwise ($g \neq h$).

If these two separate formations are used in a multiple linear regression solution, $R^2 = .49632$, $F = 4.874$, with df = 4,20 and $p < .05$ for both solutions, the same as found previously. Here, the computed t values contrast the group mean to the overall mean. Results for computed t values and correlation coefficients are the same for the usual contrast coding solution (using 1, 0 and -1) and the nonsense contrast coding solution (through different than those found for the characteristic coding scheme), although the means, regression coefficients and standard error of the regression coefficients differ from each other, as before. An equation similar to equation 8 (but even more intractable) can be developed for the nonsense contrast coding scheme.

What is the Advantages/Disadvantages of Using Nonsense Coding

Perhaps the major advantage of nonsense coding is that it should allow users of regression a larger understanding of the coding process, and the "robustness" of the coding procedures. On occasion, a particular nonsense coding scheme may make a "bit of sense" in that application. On the other hand, simple binary (1 or 0) coding is much easier to learn and to interpret the outcomes. Perhaps then the major use of nonsense coding is to instill in regression users a sense of versatility in the regression methodology.

References

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, J. D. (1974a). A note on contrast coding vs. dummy coding. *Multiple Linear Regression Viewpoints*, *4*, 1-5.

Williams, J. D. (1974b). *Regression analysis in educational research*. New York: MSS Publishing Company.

Williams, J. D. (1976). Multiple comparisons by multiple linear regression. *Multiple Linear Regression Viewpoints*, *7*(1).