In Neil Salkind (Ed.), ***Encyclopedia of Research Design.***
Thousand Oaks, CA: Sage. 2010

Effect Coding

Hervé Abdi

1 Overview

Effect coding is a coding scheme used when an analysis of variance (ANOVA) is performed with multiple linear regression (MLR). With effect coding, the experimental effect is analyzed as a set of (non-orthogonal) contrasts which opposes all but one experimental conditions to one given experimental condition (usually the last one). With effect coding, the intercept is equal to the grand mean and the slope for a contrast expresses the difference between a group and the grand mean.

2 Multiple Regression framework

In linear multiple regression analysis, the goal is to predict, knowing the measurements collected on N subjects, a dependent variable Y from a set of J independent variables denoted

$$\{X_1, \ldots, X_j, \ldots, X_J\} \ . \tag{1}$$

Hervé Abdi
The University of Texas at Dallas
Address correspondence to:
Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/∼herve

We denote by $\mathbf{X}$ the $N \times (J+1)$ augmented matrix collecting the data for the independent variables (this matrix is called augmented because the first column is composed only of ones), and by $\mathbf{y}$ the $N \times 1$ vector of observations for the dependent variable. These two matrices have the following structure.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,k} & \cdots & x_{1,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,k} & \cdots & x_{n,K} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{N,k} & \cdots & x_{N,K} \end{bmatrix} \quad \text{and } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \\ \vdots \\ y_N \end{bmatrix} \tag{2}$$

The predicted values of the dependent variable $\widehat{Y}$ are collected in a vector denoted $\widehat{\mathbf{y}}$ and are obtained as:

$$\widehat{\mathbf{y}} = \mathbf{Xb} \quad \text{with} \quad \mathbf{b} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y} \ . \tag{3}$$

the vector $\mathbf{b}$ has J component. Its first component is traditionally denoted b_0, it is called the *intercept* of the regression and it represents the regression component associated with the first column of the matrix $\mathbf{X}$. The additional J components are called *slopes* and each of them provides the amount of change in Y consecutive to an increase in one unit of its corresponding column.

The regression sum of squares is obtained as

$$SS_{\text{regression}} = \mathbf{b}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{y} - \frac{1}{N}(\mathbf{1}^{\mathsf{T}}\mathbf{y})^2 \tag{4}$$

(with $\mathbf{1}^{\mathsf{T}}$ being a row vector of 1's conformable with $\mathbf{y}$).

The total sum of squares is obtained as

$$SS_{\text{total}} = \mathbf{y}^{\mathsf{T}}\mathbf{y} - \frac{1}{N}(\mathbf{1}^{\mathsf{T}}\mathbf{y})^2 \ . \tag{5}$$

The residual (or error) sum of squares is obtained as

$$SS_{\text{error}} = \mathbf{y}^{\mathsf{T}}\mathbf{y} - \mathbf{b}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{y} \ . \tag{6}$$

The quality of the prediction is evaluated by computing the multiple coefficient of correlation denoted $R^2_{Y.1,\ldots,J}$. This coefficient is equal to the squared coefficient of correlation between the dependent variable (Y) and the predicted dependent variable ($\widehat{Y}$).

An alternative way of computing the multiple coefficient of correlation is to divide the regression sum of squares by the total sum of squares. This shows that $R^2_{Y.1,\ldots,J}$ can also be interpreted as the proportion of variance of the dependent variable explained by the independent variables. With this interpretation, the multiple coefficient of correlation is computed as

$$R^2_{Y.1,\ldots,J} = \frac{SS_{\text{regression}}}{SS_{\text{regression}} + SS_{\text{error}}} = \frac{SS_{\text{regression}}}{SS_{\text{total}}} . \tag{7}$$

2.1 Significance test

In order to assess the significance of a given $R^2_{Y.1,\ldots,J}$, we can compute an F ratio as

$$F = \frac{R^2_{Y.1,\ldots,J}}{1 - R^2_{Y.1,\ldots,J}} \times \frac{N - J - 1}{J} . \tag{8}$$

Under the usual assumptions of normality of the error and of independence of the error and the scores, this F ratio is distributed under the null hypothesis as a Fisher distribution with $\nu_1 = J$ and $\nu_2 = N - J - 1$ degrees of freedom.

3 Analysis of variance framework

For an ANOVA, the goal is to compare the means of several groups and to assess if these means are statistically different. For the sake of simplicity, we assume that each experimental group comprises the same number of observations denoted I (*i.e.,* we are analyzing a "balanced design"). So, if we have K experimental groups with a total of I observations per group, we have a total of $K \times I = N$ observations denoted $Y_{i,k}$. the first step is to compute the K experimental means denoted $M_{+,k}$ and the grand mean denoted $M_{+,+}$. The ANOVA evaluates the difference between the mean by comparing the dispersion of the experimental means to the grand mean (*i.e.,* the dispersion between means) with the dispersion of the experimental scores to the means (*i.e.*, the dispersion within the groups). Specifically, the dispersion between the means is evaluated by computing the sum of squares between means, denoted SS_{Between} and computed as:

$$SS_{\text{Between}} = I \times \sum_{k}^{K} \left(M_{+,k} - M_{+,+}\right)^2 \ . \tag{9}$$

the dispersion within the groups is evaluated by computing the sum of squares within groups, denoted SS_{Within} and computed as:

$$SS_{\text{Within}} = \sum_{k}^{K} \sum_{i}^{I} \left(Y_{i,k} - M_{+,k}\right)^2 \ . \tag{10}$$

If the dispersion of the means around the grand mean is due only to random fluctuations, then the SS_{Between} and the SS_{Within} should be commensurable. Specifically, the null hypothesis of no effect can be evaluated with an F-ratio computed as

$$F = \frac{SS_{\text{Between}}}{SS_{\text{Within}}} \times \frac{N-K}{K-1} \ . \tag{11}$$

Under the usual assumptions of normality of the error and of independence of the error and the scores, this F ratio is distributed under the null hypothesis as a Fisher distribution with $\nu_1 = K - 1$ and $\nu_2 = N - K$ degrees of freedom. If we denote by $R^2_{\text{Experimental}}$ the following ratio:

$$R^2_{\text{Experimental}} = \frac{SS_{\text{Between}}}{SS_{\text{Between}} + SS_{\text{Within}}} \ , \tag{12}$$

we can re-expressed Equation 11 in order to show its similarity with Equation 8 as:

$$F = \frac{R^2_{\text{Experimental}}}{1 - R^2_{\text{Experimental}}} \times \frac{N-K}{K-1} \ . \tag{13}$$

4 Analysis of variance with effect coding multiple linear regression

The similarity between Equations 8 for MLR and 13 for ANOVA suggests that these two methods are related and this is indeed the case.

In fact, the computations for an ANOVA can be performed with MLR via a judicious choice of the matrix $\mathbf{X}$ (the dependent variable is represented by the vector $\mathbf{y}$). In all cases, the first column of $\mathbf{X}$ will be filled with 1's and is coding for the value of the intercept. One possible choice for $\mathbf{X}$, called *mean coding*, is to have one additional column in which the value for the nth observation will be the mean of its group. This approach provides a correct value for the sums of squares but not for the F (which needs to be divided by $K-1$). Most coding schemes will use $J = K - 1$ linearly independent columns (as many columns as there are degrees of freedom for the experimental sum of squares). They all give the same correct values for the sums of squares and the F test but differ for the values of the intercept and the slopes. To implement effect coding, the first step is to select a group which is called the *contrasting group*, often this group is the last one. Then each of the remaining J groups is contrasted with the contrasting group. This is implmented by creating a vector for which all elements of the contrasting group have the value -1, all elements of the group under consideration have the value of $+1$, and all other elements have a value of 0.

With the effect coding scheme, the intercept is equal to the grand mean and each slope coefficient is equal to the difference between the grand mean and the mean of the group whose elements were coded with values of 1. This difference estimates the experimental effect of this group, hence the name of *effect* coding for this coding scheme. The mean of the contrasting group is equal to the intercept minus the sum of all the slopes.

5 effect coding: An example

The data used to illustrate effect coding are shown in Table 1. A standard ANOVA would give the results displayed in Table 2.

Table 1: A data set for an ANOVA. A total of $N = 12$ observations coming from $K = 4$ groups with $I = 3$ observations per group.

	a_1	a_2	a_3	a_4	
S_1	20	21	17	8	
S_2	17	16	16	11	
S_3	17	14	15	8	
$M_{a.}$	18	17	16	9	$M_{..} = 15$

Table 2: ANOVA table for the data from Table 1 .

Source	*df*	*SS*	*MS*	*F*	Pr(*F*)
Experimental	3	150.00	50.00	10.00	.0044
Error	8	40.00	5.00		
Total	11	190.00			

In order to perform an MLR analysis, the data from Table 1 need to be "vectorized" in order to provide the following $\mathbf{y}$ vector:

$$\mathbf{y} = \begin{bmatrix} 20 \\ 17 \\ 17 \\ 21 \\ 16 \\ 14 \\ 17 \\ 16 \\ 15 \\ 8 \\ 11 \\ 8 \end{bmatrix} . \tag{14}$$

In order to create the $N = 12$ by $J + 1 = 3 + 1 = 4$ $\mathbf{X}$ matrix, we have selected the 4th experimental group to be the contrasting group. The first column of $\mathbf{X}$ codes for the intercept, it comprises only 1s. For the other columns of $\mathbf{X}$ the values for the observations of the contrasting group will all be equal to -1. The second column of $\mathbf{X}$ will use values of 1 for the observations of the first group, the third column of $\mathbf{X}$ will use values of 1 for the observations of the second group, and the fourth column of $\mathbf{X}$ will use values of 1 for the observations of the third group:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix} . \tag{15}$$

With this effect coding scheme, we obtain the following $\mathbf{b}$ vector of regression coefficients:

$$\mathbf{b} = \begin{bmatrix} 15 \\ 3 \\ 2 \\ 1 \end{bmatrix} . \tag{16}$$

We can check that the intercept is indeed equal to the grand mean (*i.e.*, 15) and that the slopes corresponds to the difference between the corresponding groups and the grand mean. When using the MLR approach to the ANOVA, the predicted values correspond to the group means and this is indeed the case here.

6 Alternatives to effect coding

The two main alternatives to effect coding are dummy coding and contrast coding. Dummy coding is quite similar to effect coding, the only difference being that the contrasting group is always coded with values of 0 instead of -1. With dummy coding, the intercept is equal to the mean of the contrasting group and each slope is equal to the mean of the contrasting group minus the mean of the group under consideration. For contrast coding, a set of (generally orthogonal, but linear independent is sufficient) J contrasts is chosen for the last J columns of $\mathbf{X}$. The values of the intercept and slopes will depend upon the specific set of contrasts used.

Related entries

Analysis of variance, contrasts, dummy coding, mean coding, multiple linear regression.

Further readings

Abdi, H., Edelman, B., Valentin, D., & Dowling, W.J. (2009). *Experimental design and analysis for psychology.* Oxford: Oxford University Press.

Edwards, A.L.. (1985). *Multiple regression analysis and the analysis of variance and covariance*. New York: Freeman.

Fox J. (2008). *Applied regression analysis and generalized linear models*. Thousand Oaks: Sage.

Edwards, A.L.. (1985). *Data analysis for research designs*. New York: Freeman.