DOCUMENT RESUME

ED 061 276 TM 001 211

AUTHOR Lewis, Ernest L.; Mouw, John T.
TITLE The Use of Contrast Coding to Simplify ANOVA and ANCOVA Procedures in Multiple Linear Regression.
PUB DATE Apr 72
NOTE 17p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Covariance; *Analysis of Variance; Hypothesis Testing; Interaction; *Mathematical Models; *Multiple Regression Analysis; Tests of Significance
IDENTIFIERS *Contrast Coding

ABSTRACT

This paper discusses the use of contrast coefficients in multiple linear regression models, and shows how they can provide for a logical method of analysis in both the analysis of variance and the analysis of covariance. (CK)

Paper Presented to a Meeting of The American Educational Research Association

April 1972

The Use of Contrast Coding to Simplify ANOVA and ANCOVA Procedures in Multiple Linear Regression

Ernest L. Lewis
Southern Illinois University
at Carbondale

and

John T. Mouw
Southern Illinois University
at Carbondale

The Use of Contrast Coding to Simplify ANOVA and ANCOVA Procedures in Multiple Linear Regression

Cohen (1968) presented a discussion of contrast coding in multiple linear regression models for use in analysis of variance (ANOVA) and analysis of covariance (ANCOVA). The general theme of Cohen's article was that the main effects and interaction of ANOVA and ANCOVA can be reflected in a linear model through the use of specifically coded predictor vectors. Other writers have referred to these vectors as dummy vectors, nonsense coded vectors, or group membership vectors. In our work with multiple regression, we have found Cohen's system of contrast coding to provide a very logical and relatively simple method for developing regression models to answer more specific questions than the overall main effects and interaction tests generally applied in ANOVA. One purpose of this paper is to present a discussion of the use of contrast coding to reflect orthogonal comparisons.

We have also found that, as Cohen suggests, contrast coding can easily be applied in ANCOVA. Further, we found that for a two-way analysis of covariance, contrast coding leads to a more exact duplication of traditional analysis of covariance than does the standard method of designating group membership predictor vectors. A second purpose of this paper is to present a discussion of the application of contrast coding to ANCOVA.

Analysis of Variance

Consider an experiment in which two treatment conditions are to be compared. In this case, Winer (1962) indicates that each individual score results from a number of sources of variability. According to Winer,

$$X_{ij} = \mu + \tau_j + e_{ij}$$

Where: X_{ij} = an observation on person i under treatment j

μ = grand mean of all potential observations

τ_j = effect of treatment j

e_{ij} = error associated with X_{ij}

In order to answer the question of whether there is a significant difference between Treatments 1 and 2 in a standard regression model (Bottenberg & Ward, 1963; Kelly, et.al., 1969), one would employ the following full model:

Model 1 $Y = a_0U + a_1X_1 + a_2X_2 + E_1$

Where: Y = vector of criterion scores

U = unit vector (all elements are 1)

X_1 = 1 if the corresponding criterion score comes from Treatment 1; 0 otherwi

X_2 = 1 if the corresponding criterion score comes from Treatment 2; 0 otherwise

E_1 = error vector

a_0, a_1, a_2 = partial regression weights

It will be noted that Y corresponds to Winer's X_{ij}; a_0 to Winer's μ; a_1 and a_2 to Winer's τ_j and E_1 to Winer's e_{ij}. To determine if [illegible] ant difference exists between Treatments 1 and 2, Model 1 would be compared to a restricted model (Model 99) which would contain only the unit vector as a predictor vector and the error vector.

Model 99 $Y = a_0U + E$

Using contrast coding to reflect Treatments 1 and 2, the following full model would result:

Model 2 $\quad Y = a_0U + a_1X_1 + E_2$

Where: $\quad Y$ = criterion scores

U = unit vector

X_1 = 1 if criterion from Treatment 1 or -1 if criterion from Treatment 2

E_2 = error vector

a_0, a_1 = partial regression weights

To answer the question as to whether or not Treatments 1 and 2 are different, Model 2 would be compared to Model 99.

The advantage of contrast coding in the above example seems to be in the determination of degrees of freedom. It will be noted that the analysis in this example consists of a simple t-test or an F-test with one degree of freedom in the numerator. In order to perform this analysis, one must set a_1 and a_2 from Model 1 equal to 0. This loss of two vectors results in a loss of only one degree of freedom because there is a linear dependency existing within the set of vectors U, X_1, and X_2 in Model 1. In Model 2, no linear dependencies exist in the predictor variables. As a result, the test for a significant difference between Treatments 1 and 2 is accomplished simply by setting $a_1 = 0$. As a result, the restriction of one regression weight accurately reflects the appropriate number of degrees of freedom for this analysis.

If one were to expand the above two-group example to include four treatment conditions, the advantages of contrast coding in ANOVA become more apparent. If Treatments 3 and 4 are added, the addition of X_3 and X_4 to Model 1 would be required in order to allow for the main effects of Treatments 3 and 4. Model 1 would then be revised to be:

Model 3 $Y = a_0U + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + E_3$

Where: Y, U, X_1, X_2 and E_3 are as defined in Model 1

$X_3 = 1$ if T_3, 0 otherwise

$X_4 = 1$ if T_4, 0 otherwise

a_0, a_1, a_2, a_3, a_4 = partial regression weights

To test for an overall main effect of treatments, the following restriction would be placed on Model 3:

$$a_1 = a_2 = a_3 = a_4 = 0$$

It can again be seen that there is one more predictor vector restricted out than degrees of freedom lost. Further, it should be noted that within this regression model framework, the overall treatment main effects is the only question which can be asked and tested.

Using contrast coding, Model 4 might be used to reflect the various treatment conditions.

Model 4 $Y = a_0U + a_1X_1 + a_2X_2 + a_3X_3 + E_4$

Where: Y = criterion scores

U = unit vector

E_4 = error vector

a_0, a_1, a_2, a_3 = partial regression weights

and where the elements in X_1, X_2, and X_3 reflect the linear, quadratic and cubic trends and are as follows:

If criterion from Treatment	X_1	X_2	X_3
1	-3	1	-1
2	-1	-1	3
3	1	-1	-3
4	3	1	1

The elements presented here are the standard coefficients for orthogonal polynomials. The use of these values would result in X_1, X_2, and X_3 being uncorrelated. As a result, it is possible to partition the variance into its three independent sources.

If one were concerned about asking the overall main effect question, it would be necessary to set $a_1 = a_2 = a_3 = 0$. The test of significance would result in precisely the same outcome as the use of 1 and 0 group membership vectors as presented in Model 3. However, it is possible to ask more specific questions given orthogonal coefficients. One may not only be interested in the overall main effect question. The research hypothesis in a particular research project might be that "the average of Treatment Groups 1 and 4 is different from the average of Treatment Groups 2 and 3" (in other words, whether the difference follows a quadratic trend). Given Model 4, it would simply require the a_2 be set equal to 0 in order to answer this very specific question.

As indicated above, the values in the vectors are standard coefficients for orthogonal polynomials. It may be that such coefficients do not reflect a particular question of interest. One might want to ask the question as to whether the effect of Treatment 1 equals the average effect of Treatments 2, 3, and 4. Since the standard coefficients for orthogonal polynomials do not reflect this particular question, it would be necessary to establish a different set of coding coefficients. Since the question as to whether Treatment 1 equals the average of Treatments 2, 3 and 4 would require coding coefficients in the predictor vectors to reflect the differential weighting of the Treatments, an appropriate set of coding coefficients might be: Treatment 1 = 3; Treatment 2 = -1; Treatment 3 = -1; Treatment 4 = -1. The values in vectors X_1, X_2, and X_3 of Model 4 might then be as follows:

	X_1	X_2	X_3
If criterion from Treatment			
1	3	0	0
2	-1	2	0
3	-1	-1	+1
4	-1	-1	-1

In order to answer this question of interest, it would only be necessary to restrict out vector X_1 by setting $a_1 = 0$.

The two examples presented above seem to point to two advantages which accrue from the use of contrast coding in a one-way analysis of variance. First, since the predictor vectors are all independent, the number of predictor variables in a model accurately reflects the degrees of freedom for the analysis. As was pointed out above, this is not the case when standard 1 and 0 group membership vectors are used. Second, the use of contrast coding allows one to ask more specific questions of interest than the overall main effect. The importance of these two factors becomes even more apparent when one considers a two-way analysis of variance.

Consider an experiment in which a 2 x 3 factorial design is to be applied and assume that there are two levels of condition A and three levels of condition B. Winer indicates that the following linear model would account for all sources of variability contributing to an individual score:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

Where: X_{ijk} = an observation on person k under treatment i and treatment j

μ = grand mean of all potential observations

α_i = main effect for condition A

β_j = main effect for condition B

$\alpha\beta_{ij}$ = effect of interaction of conditions A and B

e_{ijk} = error associated with X_{ijk}

The various sources of variability in Winer's model can be duplicated in standard multiple regression analysis. However, the need for a test of interaction requires a full model which allows the differences between cell means to vary and a restricted model which would force the differences between cell means to be equal. While this is not a particularly difficult task, it does require some rather lengthy algebraic manipulations of the partial regression weights. Kelly, Beggs, McNeil, Eichelberger and Lyon (1969) include an excellent presentation of the procedures for performing a two-way analysis of variance in standard regression analysis so we will not attempt to duplicate it here.

Using contrast coding to duplicate the 2 x 3 analysis of variance would require the following full model:

Model 5 $Y = a_0U + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + E_5$

Where: Y = criterion vector

U = unit vector

X_1 = 1 if subject from A_1 or -1 if from subject A_2

X_2 = -1 if subject from B_1; 0 if subject from B_2 or 1 if subject from B_3.

X_3 = 1 if subject from B_1; -2 if subject from B_2 or 1 if subject from B_3.

X_4 = X_1 multiplied by X_2

X_5 = X_1 multiplied by X_3

E_5 = error vector

a_0, through a_5 = partial regression weights

It will be noted that the elements of X_2 and X_3 reflect the linear and quadratic trends for the B main effect. In addition, the coefficients in X_4 and X_5 would reflect the linear and quadratic components of the interaction effect. It should also be noted that all five vectors are independent so that the number of predictor vectors accurately reflects the total degrees of freedom for this two-way analysis of variance.

The overall main effects and interaction tests can be simply done once Model 5 has been established. In order to test for interaction, one need only set $a_4 = a_5 = 0$ and compare the R^2 of Model 5 to the R^2 of the resulting restricted model. In order to test for a significant A main effect one need only restrict X_1 from Model 5 by setting $a_1 = 0$. To test for the B main effect, X_2 and X_3 must be restricted from Model 5 by setting $a_2 = a_3 = 0$. Each of these tests of significance can be shown to exactly duplicate the results one would obtain through the use of traditional two-way analysis of variance equations.

As was the case with a one-way analysis of variance, the use of contrast coefficients allows one to ask questions of interest other than the overall main effects and interaction. In the example above, suppose one were interested in determining if the interaction contained a significant quadratic trend. This variable of interest is reflected in X_5 of Model 5. In order to test for a significant quadratic interaction trend, one need only set $a_5 = 0$. The linear trend of the interaction could be tested by setting $a_4 = 0$. Further, Model 5 allows one to test for significant linear and quadratic components of the B main effect by setting $a_2 = 0$ and $a_3 = 0$ respectively. The use of contrast coefficients in this linear regression analysis would allow one to examine any one or all of the five independent sources of variance which the total degrees of freedom indicate contribute to each individual criterion score. In addition, one could ask other questions of interest by establishing a set of contrast codes which would allow the specific question of interest to be reflected in the predictor vectors.

Analysis of Covariance

The application of the use of contrast coefficients for analysis of covariance is a natural extension of the analysis of variance. The covariate or concomitant variable is entered as a predictor along with the treatment variables in the linear equation. For example, if a covariate were included in Model 5 above the equation would become:

Model 6 $Y = a_0U + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + E_6$

Where: Y = criterion scores

X_1 through X_3 are treatment variables of interest

X_4 = the covariate or concomitant variable

U = unit vector

E_6 = error vector

a_0 through a_4 = partial regression weights

The nature of the equation changes slightly, however, in that the predictor variables (X_1 through X_4) are not all orthogonal to one another. Specifically, there is a real or sample covariance between the covariate (X_4) and each of the variables of interest (X_1 through X_3). When the restriction $a_1 = a_2 = a_3 = 0$ is placed on the equation eliminating the treatment source of variance, the weight associated with the covariate (a_4) will change in value. It can be shown that the variance which is lost by such a restriction is that variance which is associated with the treatment but which is independent of the covariate. In other words, the restriction results in a loss of that variance which is unique to the treatment variables (X_1 through X_3). Such analysis is identical to analysis of covariance as described in such textbooks as Winer (1962), Lindquist (1953) and McNemar (1969). The interpretation made for a significant statistical test for treatment effect obtained by the analysis is that the treatments have an effect on the mean criterion scores over and above that which is accounted for by the covariate. The usual procedure of using group membership vectors in the linear model also duplicates the analysis of covariance for one-way ANCOVA designs. In fact, the only advantages for using contrast coefficients rather than group membership vectors seem to be that (1) contrast coefficients provide a more direct count of independent vectors to obtain degrees of freedom and (2) contrast coefficients allow for tests of more specific questions concerning treatment effects than does the use of group membership vectors.

Winer (1962) indicates that the linear model for a two-factor ANCOVA would be as follows:

$$X_{ijk} = \mu + \alpha_i + \beta_{ij} + \alpha\beta_{ij} + \gamma_{yk} + e_{ijk}$$

Where: X_{ijk} = an observation on person k under treatment i in condition j given information on the covariate

μ = grand mean of all observations

α_i = effect of the ith treatment

β_j = effect of the jth treatment

$\alpha\beta_{ij}$ = effect due to interaction

γ_{vk} = regression effect on the covariate

e_{ijk} = error associated with X_{ijk}

Suppose, now, that we wish to utilize a model where the α_i effect contains two different conditions and the β_j effect consists of a control group (B_1) and two experimental groups (B_2 and B_3). Then the more traditional regression model for these effects with the covariate and interaction included would be:

Model 7 $\quad Y = a_0U + a_7A_1B_1 + a_8A_1B_2 + a_9A_1B_3 + a_{10}A_2B_1 + a_{11}A_2B_2 + a_6X_0 + E_7$

Where: Y = vector of criterion scores

U = unit vector (all elements are 1)

A_iB_j = 1 if observation is found in both A_i and B_j, 0 otherwise

X_0 = concomitant variable

a_0 and a_6 through a_{12} = partial regression weights

In order to test the interaction effect, one would restrict Model 7 to:

Model 8 $Y = a_0U + a_1A_1 + a_2A_2 + a_3B_1 + a_4B_2 + a_5B_3 + a_6X_0 + E_8$

Where: Y = vector of criterion scores

U = unit vector

A_1 = 1 if criterion from A_1; 0 otherwise

A_2 = 1 if criterion from A_2; 0 otherwise

B_1 = 1 if criterion from B_1; 0 otherwise

B_2 = 1 if criterion from B_2; 0 otherwise

B_3 = 1 if criterion from B_3; 0 otherwise

X_0 = concomitant variable

E_8 = error vector

a_0 through a_6 = partial regression weights

Then the test for interaction ($R_7^2 - R_8^2$) would be a test of whether the proportion of variance unique to interaction is significant.

There is some disagreement among researchers as to procedures for testing main effects following a non-significant test for interaction. Both Ferguson (1971) and Winer (1962) suggest that after finding a non-significant interaction effect, one has the option of treating the interaction sums of squares as error. The sums of squares for interaction could, along with the appropriate degrees of freedom, be pooled with the sums of squares error to form a more stable error estimate. In another paper presented at this convention, Pohlmann (1972) discusses the limit to which such pooling may aid in guarding against a type II error.

Kelly et.al. (1969) encourage the practice of pooling as discussed in the previous paragraph. Assuming one has chosen to pool, then the A effect could be tested by restricting a_1 and a_2 from Model 8 equal to 0 and the subsequent model becomes:

Model 9 $Y = a_0U + a_3B_1 + a_4B_2 + a_5B_3 + a_6X_0 + E_9$

Where: Y = vector of criterion scores

U = unit vector

E_9 = error vector

B_1, B_2, and B_3 = defined as in Model 8

X_0 = concomitant variable

a_0, a_3, a_4, a_5, a_6 = partial regression weights

$(R_8^2 - R_9^2)$ would seem to be equal to the A main effect whereas $1 - R_8^2$ would consist of a pooled error term which includes the interaction effect. The B main effect would be tested in a manner similar to the test for the A effect.

This, however, does not duplicate the main effect that is found in traditional two factor ANCOVA as described in Winer (1962). In the model:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \gamma_{yk} + e_{ijk}$$

α_i, β_j, $\alpha\beta_{ij}$ are orthogonal to one another and, hence, the presence or absence of any one should not have any effect on the others. However, this is not the case in the presence of the covariate. The covariance patterns between α_i, β_j, and $\alpha\beta_{ij}$, with the covariate γ seem to be of such a nature that the restriction of any of the three effects equal to 0 results in a change (increase or decrease) in the other remaining effects. This would not be the case without the presence of the covariate nor does it effect a one-way ANCOVA. Thus, when the interaction term is pooled with the error in order to test a main effect, the amount of variance associated with that main effect is different from what it would have been without pooling.

The use of contrast coding in two-factor ANCOVA would provide a method of analysis where one could very easily test the main effects without pooling the interaction, thus yielding a duplicate result to the traditional two-factor ANCOVA as discussed by Winer (1962). Furthermore, contrast coefficients allow for tests of more specific questions of interest.

Given the example presented above, where the A effect consists of two conditions and the B effect consists of one control group (B_1) and two experimental groups (B_2 and B_3), the experimenter might be interested in a comparison of the experimental groups of the B condition to the control group (B_1) as the first question of interest. A second question might be if there is a difference between the experimental groups over both A conditions. Then, one may be interested in whether either or both of the B experimental effects are different within the two A conditions.

Note that in all questions, the interest lies in the effect of treatment over and above that of the concomitant variable.

The model appropriate for this ANCOVA would be as follows:

Model 10 $Y = a_0U + a_1A + a_2B_1 + a_3B_2 + a_4AB_1 + a_5AB_2 + a_6X_0 + E_{10}$

Where: Y = criterion vector

U = unit vector

A = 1 if in condition A_1; -1 if in condition A_2

B_1 = 2 if in control group; -1 if in either experimental group

B_2 = 0 if in control group; 1 if in experimental group 1 (B_2); -1 if in experimental group 2 (B_3)

AB_1 = (obtained by A x B_1) $A_1B_1 = 2$; $A_1B_2 = -1$; $A_1B_3 = -1$;

$A_2B_1 = -2$; $A_2B_2 = 1$; $A_2B_3 = 1$

AB_2 = (obtained by A x B_2) $A_1B_1 = 0$; $A_1B_2 = 1$; $A_1B_3 = -1$;

$A_2B_1 = 0$; $A_2B_2 = -1$; $A_2B_3 = 1$

X_0 = concomitant variable

E_{10} = error vector

a_0 through a_6 = the regression weight associated with the respective vectors

There are three apparent advantages of contrast coding over the more standard use of group membership vectors. First, the number of parameter estimates are directly reflected by the number of weights (a_0 through a_6) used in the model and, hence, lead to a more direct count of degrees of freedom. Secondly, one can go directly to tests of the questions of interest by restricting the appropriate weight of Model 10. For the four questions of interest specified above this would result in four restricted models by first setting $a_2 = 0$, followed by $a_3 = 0$, $a_4 = 0$, and finally $a_5 = 0$. Each of the resulting restricted models would be compared to Model 10 above. One could still test for an overall interaction effect or for either of the main effects (A or B), by simply restricting all weights for the appropriate vectors equal to 0. The third advantage of contrast coding is that it allows for tests of the main effect without pooling the error term, thus precisely duplicating the two factor ANCOVA as presented in Winer (1962). While such a traditional analysis may not be superior, the authors suspect that the difference in the two analyses would lead to somewhat different conclusions. That is, the traditional ANCOVA and the use of contrast coefficients analyze variance which is independent of all other sources in the model; whereas, the use of standard group membership vectors yields variance components that are in some way common to the interaction.

Summary

In this paper, we have shown how the use of contrast coefficients in multiple linear regression models can provide for a logical method of analysis in ANOVA and ANCOVA. Three distinct advantages were indicated. First, the number of estimated parameters are directly indicated in the model, thus leading to a more natural and direct count for degrees of freedom. Second, contrast coding allows for the testing of specific variables of interest other than the overall main effect and overall interaction effects. Finally, in the case of two-way ANCOVA, contrast coding does not require pooling interaction with the error term and thus is an exact duplicate of ANCOVA as presented in Winer (1962).

It would seem that the use of contrast coefficients allow for a variety of types of analysis within the general linear model. This would present future researchers with a more integrated concept of data analysis rather than to contribute to fragmentation of the field by discussing regression as separate from ANOVA with all its various subcategories. The use of contrast coefficients encourages researchers to ask specific questions which can be analyzed with F-tests which have only one degree of freedom in the numerator. When there is only one degree of freedom in the numerator, the researcher is in effect dealing with a single source of variability, and as a result, is able to better interpret the meaning of the test of significance. In overall main effects or interaction tests, the numerator generally has more than one degree of freedom in the numerator. The researcher must then attempt to interpret the test of significance realizing that he is analyzing several sources of variability simultaneously.

References

Bottenberg, R.A., & Ward, J.H., Jr. Applied multiple linear regression. (PRL-TDR-63-6), Lackland Air Force Base, Texas, 1963.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, Vol. 70, 6, 426-443, 1968.

Kelly, F.J., Beggs, D.L., McNeil, K.A., Eichelberger, R.T., & Lyon, J. Research design in the behavioral sciences: multiple regression approach. Southern Illinois University Press, Carbondale and Edwardsville, Illinois, 1969.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Houghton Mifflin, Boston, 1953.

McNemar, O. Psychological statistics. John Wiley & Sons, New York, 1969.

Pohlmann, J. The effects of pooling the interaction and within components on the alpha and power of the main effects test. Unpublished paper presented at AERA, 1972.

Winer, B.J. Statistical principles in experimental design. McGraw-Hill, New York, 1962.