**SPSS Library  
Understanding and Interpreting Parameter Estimates in Regression and ANOVA**

**This page was adapted from a web page at the** [**SPSS web page**](http://www.spss.com)**.  We thank SPSS for their permission to adapt and distribute this page via our web site.**

**This page is composed of 5 articles from SPSS Keywords exploring issues in the understanding and interpretation of parameter estimates in regression models and anova models.  The figures have been renumbered consecutively.**

**You can download the data file ( [catreg.sav](http://www.ats.ucla.edu/stat/spss/library/catreg.sav) ) used in this article so you can reproduce the results for yourself.**

**USING CATEGORICAL VARIABLES IN REGRESSION  
David P. Nichols  
Senior Support Statistician  
SPSS, Inc.  
From SPSS Keywords, Number 56, 1995**

When we polled Keywords readers to find out what kinds of topics they most wanted to see covered in future Statistically Speaking articles, we found that many SPSS users are concerned about the proper use of categorical predictor variables in regression models. Since the interpretation of the estimated coefficients is a major part of the analysis of a regression model, and since this interpretation depends upon how the predictors have been coded (or in technical terms, how the model has been parameterized), this is indeed an important topic.

To begin with, we will assume that the model under consideration involves only first order or main effects of predictor variables. That is, no higher order polynomial terms such as squares or cubes are used, and no interactions between predictors are involved. Such higher order or product terms introduce complexities beyond those introduced by the presence of main effects involving categorical variables. We will avoid these complexities for the time being. We will further assume that we have complete data; that is, no missing values on any predictor or dependent variables. We begin with a brief review of the interpretation of estimated regression coefficients.

As you may remember, in a linear regression model the estimated raw or unstandardized regression coefficient for a predictor variable (referred to as B on the SPSS REGRESSION output) is interpreted as the change in the predicted value of the dependent variable for a one unit increase in the predictor variable. Thus a B coefficient of 1.0 would indicate that for every unit increase in the predictor, the predicted value of the dependent variable also increases by one unit. In the common case where there are two or more correlated predictors in the model, the B coefficient is known as a partial regression coefficient, and it represents the predicted change in the dependent variable when that predictor is increased by one unit while holding all other predictors constant. The intercept or constant term gives the predicted value of the dependent variable when all predictors are set to 0.

For our purposes the important distinction between different types of predictor variables is between those measured on at least an interval scale, where a change of one unit in the predictor has a constant meaning across the entire scale, and those where such consistency of unit differences is not assumed. Though these are theoretically distinct, in practice it often happens that the terms interval and subinterval are replaced by continuous and categorical. The interpretation of estimated regression coefficients given above applies in a fairly straightforward manner to interval predictors, continuous or not, and their use in procedures like REGRESSION is quite simple as a practical matter: just name them as independent variables and specify when you want them used. For subinterval variables, which is the assumption in SPSS for categorical variables, things are more complicated. Despite the fact that equating continuous with interval and categorical with subinterval is an abuse of language, we'll proceed to do just that, to avoid confusion related to use of SPSS procedures.

One reason that the handling of categorical predictors is so important is that by the time one gets to the actual computation of the regression equation, no distinction is made between subinterval and interval variables. To put it another way, a matrix algebra routine knows nothing about different types of numbers; they're all just numbers. Some SPSS procedures used to analyze linear and generalized linear regression models are designed to handle the translation from categorical to interval representations with only minimal guidance from the user. These include the T-TEST procedure, the analysis of variance procedures ONEWAY, ANOVA and MANOVA, and the newer nonlinear regression procedures LOGISTIC REGRESSION and COX REGRESSION. However, even when such automatic handling of categorical predictors is available, it is still incumbent upon the user to make sure that he or she understands categorical variable representations well enough to produce useful results and to be able to interpret these results.

The simplest possible regression involving categorical predictors is one with a single dichotomous (two level) independent variable. An example of such a regression model would be the prediction of 1990 murder rates in each of the 50 states in the U.S.A. based upon whether or not each state had a death penalty statute in force just prior to and during that time. The data are compiled from almanac sources; murder rates are measured in number per 100,000 population. The variable of interest, denoted MURDER90, has a mean value of about 4.97 for the fourteen states without a death penalty statute, and about 7.86 for the 36 states with the death penalty.

Figure 1 presents the results of a dummy variable regression of MURDER90 on DEATHPEN, a categorical variable taking on a value of 0 for the no death penalty states and 1 for the death penalty states. 0-1 coding, known as dummy or indicator coding, is quite popular, as it often lends itself to the simplest possible interpretation.

**Figure 1**

**---------------------------------------------------------------------------**

**Multiple R .33556**

**R Square .11260**

**Adjusted R Square .09411**

**Standard Error 3.72103**

**Analysis of Variance**

**DF Sum of Squares Mean Square**

**Regression 1 84.33257 84.33257**

**Residual 48 664.61163 13.84608**

**F = 6.09072 Signif F = .0172**

**------------------ Variables in the Equation ------------------**

**Variable B SE B Beta T Sig T**

**DEATHPEN 2.892460 1.172015 .335562 2.468 .0172**

**(Constant) 4.971429 .994488 4.999 .0000**

**---------------------------------------------------------------------------**

Here we have two coefficients, a constant or intercept term, and a "slope" coefficient for the DEATHPEN variable. Recall that the interpretation is that the constant is the predicted value when all predictors are set to 0, which here simply represents those states with no death penalty. Thus the constant coefficient is equal to the mean murder rate for this group. The DEATHPEN coefficient is the predicted increase in murder rate for a unit increase in the DEATHPEN variable. Since those states with a DEATHPEN value of 1 are those states with a death penalty statute, this coefficient represents the change in estimated or predicted murder rate for these states relative to those without the death penalty. The 2.89 value is exactly the difference between the two means, so that adding it to the constant produces the mean for the death penalty states. Since we are considering the entire population of states, the significance level is not necessary of particular interest, though if we were to conceptualize the current situation as resulting from a sampling from some hypothetical populations, the p-value of .0172 would indicate that so large a coefficient is unlikely to result from chance were random samples of this size drawn from hypothetical populations with equal means.

Other results of note are that the p-value for the t-test for the MURDER90 coefficient is the same as that for the overall regression F-test. This is due to the fact that the t-test tests the null hypothesis that this coefficient is 0 in the population, while the F-test tests the null hypothesis that all coefficients other than intercept are 0 in the population, and with only one predictor, these hypotheses are the same. The F-value is precisely the square of the t-value. This holds only for a simple regression involving one predictor. Also of note is the fact that the Multiple R, which reduces to the absolute value of the correlation between the predictor and the dependent variable in a simple regression, is equal to the standardized regression coefficient (Beta). In a simple regression, the standardized coefficient is the correlation between the predictor and dependent variables, and is thus constrained to be between -1 and +1. Note that this generally holds true only for a simple regression, and that with correlated predictor variables, the standardized coefficients may be larger than 1 in absolute value.

This correlation between a dichotomous variable and a continuous variable is sometimes known as a point-biserial correlation. No special formula is required; special computational formulas in texts are simply special cases of the general Pearson product moment correlation coefficient formula applied to this combination of variable types. If both variables are dichotomous, the standard formula reduces further to that for a phi coefficient.

Finally, note that there are a number of ways in SPSS to achieve the same results we obtained from REGRESSION, if our purpose were to test the null hypothesis of equality of means between the two groups of states drawn from our hypothetical populations. Precisely the same t-statistic (or the negative of the value from REGRESSION, which means the same thing, given the variable codings) could be obtained from T-TEST, the CONTRAST option in ONEWAY or parameter estimate output in MANOVA, and the F-statistic could be duplicated in ONEWAY, ANOVA or MANOVA. In ONEWAY or ANOVA we would have to use the dummy variable for DEATHPEN as a two level factor, while in MANOVA we could either specify it as a factor or as a covariate. The results in any case would be the same in terms of test statistics and p-values. One example is given in Figure 2, using default DEVIATION contrasts in MANOVA:

**Figure 2**

**---------------------------------------------------------------------------**

**Tests of Significance for MURDER90 using UNIQUE sums of squares**

**Source of Variation SS DF MS F Sig of F**

**WITHIN+RESIDUAL 664.61 48 13.85**

**DEATHPEN 84.33 1 84.33 6.09 .017**

**- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -**

**Estimates for MURDER90**

**--- Individual univariate .9500 confidence intervals**

**CONSTANT**

**Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper**

**1 6.41765873 .58601 10.95150 .00000 5.23941 7.59591**

**DEATHPEN**

**Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper**

**2 -1.4462302 .58601 -2.46794 .01721 -2.62448 -.26798**

**---------------------------------------------------------------------------**

We see that the significance level for the t-test for the DEATHPEN parameter estimate is the same as we obtained in REGRESSION. However, it is opposite in sign to our earlier coefficient, and only half the size. Our constant has also changed; note that it's value is now halfway between the means of the two groups. The differences we see here are due to the use of a different set of predictor codings being used internally by MANOVA. That is, MANOVA has parameterized the model somewhat differently than we did earlier. The default DEVIATION contrasts in MANOVA are designed to compare each level of a factor to the mean of all levels. In this case the DEATHPEN coefficient compares the no death penalty mean to the simple average of the two group means. The CONSTANT coefficient is this simple mean of group means. The F-statistic remains the same, the square of the t-value, as was the case in REGRESSION.

These results point to three important features of the regression model. One is that the interpretation of the estimated model coefficients is dependent upon the parameterization of the model; in order to know how to interpret the coefficients, we must be aware of how the predictor values have been coded. Second, despite having used two different parameterizations, we obtained the same results in terms of the test statistics for DEATHPEN. This result would occur regardless of the two numerical values used to represent the groups; all we could do by changing these values would be to flip the sign of the coefficient and inflate or deflate the coefficient and it's standard error by an equal factor, so that the absolute value of the ratio remains the same. This is true because the existence of only two groups means that a numerical representation of a comparison between them can only be done in one way; any differences in practical results are due to scaling considerations. Another way of saying this is to note that since we have only two groups, there can be only one degree of freedom in any test used to compare them, and the results must therefore always be the same. Finally, though the identical error sums of squares only intimate this and do not necessarily prove it, it is the case that the predicted values produced by the two approaches are identical. In other words, we have really fitted the same overall model in two slightly different ways.

We have yet to identify the codings given to the two levels of DEATHPEN that resulted in the MANOVA parameter estimates. In MANOVA we specified DEATHPEN as a categorical factor variable with codes of 0 and 1, and had the procedure internally create the design or basis matrix required for the model fitting. In REGRESSION, only the constant or intercept column of 1's is provided automatically by the procedure; the other columns are provided by the user in the form of the predictor variables specified. In MANOVA, the procedure automatically creates a set of predictor variables to represent a factor instead of requiring the user to do so. In the case of a dichotomous factor, MANOVA creates only one predictor in addition to the constant term, and by default it gives this variable values of 1 and -1, respectively. In our example, the states without the death penalty are the first group (having factor variable value 0), and are coded 1, while states with the death penalty receive a value of -1.

If we recall the interpretation of the regression coefficient as the increase in the predicted value of the dependent variable for a unit increase in the predictor, we can see why the DEATHPEN coefficient in MANOVA is -1/2 that of the one in REGRESSION. First, the directionality has been changed. That is, an increase in the predictor means moving from the death penalty group toward the no death penalty group. Thus the change in sign. Second, in order to compare the two groups in this parameterization, we must move two units, from -1 to 1, rather than from 0 to 1. Thus the two parameterizations are really telling us exactly the same thing. This is further illustrated by using the MANOVA results to predict the murder rates of the two groups. For the states with no death penalty, we add the CONSTANT and DEATHPEN coefficients, giving us a predicted value of about 4.97. For the death penalty group, we subtract the DEATHPEN coefficient from the CONSTANT, and obtain a predicted value of about 7.86. These are of course the same values obtained using REGRESSION.

What if we wanted to produce the same estimates in MANOVA that we obtained in REGRESSION? The only straightforward way to produce exactly the same estimates would be to enter the DEATHPEN predictor as a covariate coded 0-1. (There is a way to trick MANOVA into providing the same coefficients as REGRESSION even with DEATHPEN as a factor, but we'll ignore that here.) The reason for this is that in it's automatic reparameterization or internal recoding of the factor(s), MANOVA enforces a sum to 0 restriction on the values of the category codings. Thus 0-1 coding is not available. We can still obtain the same parameter value for the difference between the two groups of states however. This can be obtained by using SIMPLE contrasts with the first category as the reference category. This uses category codes of -1/2 and 1/2, so that an increase of one unit in the predictor would mean a change from no death penalty to the death penalty, and the resulting coefficient would be the same in both magnitude and sign as that given in REGRESSION. However, the constant or intercept term would still be the unweighted mean of the two group means. Using the CONSTANT coefficient from the MANOVA output plus or minus 1/2 times the DEATHPEN coefficient from the REGRESSION output, you can verify that this parameterization again produces exactly the same predicted values as our earlier approaches.

So much for the simple situation of a dichotomous predictor. As we have seen, in this situation the coding of the variable is important in interpreting the value of the regression coefficient, but not when we want to test whether the predictor has a nonzero population relationship with the dependent variable. One way to think about this fact is that when there are only two values of a predictor, there is only one interval between those values, so the assumption of equal meanings of intervals is automatically satisfied. However, once we move to predictors with more than two levels, things become more complicated. We'll save those complications for the next issue.

**USING POLYTOMOUS PREDICTORS IN REGRESSION  
David P. Nichols   
Senior Support Statistician   
SPSS, Inc.   
From SPSS Keywords, Number 57, 1995**

Here we extend our discussion begun last issue on categorical predictors in linear regression to the case of a three level predictor variable. One way to do this in our example is to distinguish between states with a death penalty statute but no executions in a relevant time period and those where one or more executions took place during that time. Here we will use a new variable called STATUS89, which takes on a value of 0 for those states with no death penalty, 1 for the 28 states with death penalty statutes in force but no executions taking place in 1989, and 2 for the eight states where executions occurred during 1989. The no death penalty states are the same 14 states from last time, with a mean 1990 murder rate of about 4.97. The 28 states with statutes but no 1989 executions had a rate of about 6.98, and the eight executing states had a rate of about 10.96.

How should we represent our new three level predictor? The dummy variable approach is again probably the most popular one among applied researchers. Since there are three groups, we have two degrees of freedom for making comparisons. Thus we need two dummy variables to represent STATUS89. If we make STATUS1 a dummy variable with a value of 1 for the states with a value of 1 for STATUS89 and 0 otherwise, and make STATUS2 a dummy variable with a value of 1 for the states with a value of 2 for STATUS89 and 0 otherwise, then the STATUS1 coefficient will compare states with a value of 1 on STATUS89 with those having a value of 0, and the STATUS2 coefficient will compare states with a value of 2 on STATUS89 with those having a 0 value. One useful way to look at what we are doing is to lay out the design matrix we are using at the cell level; that is, to list out the values of the new predictor variables once for each different level of the original predictor. This listing is given in Figure 3:

**Figure 3**

**---------------------------------------------------------------------------**

**STATUS89 CONSTANT STATUS1 STATUS2**

**0 1 0 0**

**1 1 1 0**

**2 1 0 1**

**---------------------------------------------------------------------------**

One question that may come to mind here is how does STATUS1 compare level 1 of STATUS89 to level 0 and not to both levels 0 and 2? Similarly, how does STATUS2 represent a comparison between level 2 of STATUS89 and level 0 and not between level 2 and the other two levels? To see why this is the case, note that cases at level 1 of STATUS89 differ from those at level 0 only on STATUS1, while cases at level 2 differ from those at level 0 only on STATUS2. From this coding layout, we can see that the states at level 0 of STATUS89 are predicted by the CONSTANT, while states at level 1 are predicted by the sum of the CONSTANT and STATUS1 coefficients, and the states at level 2 of STATUS89 are predicted by the sum of the CONSTANT and STATUS2 coefficients. Thus our regression coefficients should yield the mean for the no death penalty states as the CONSTANT, the difference between the non-executing death penalty states and the no death penalty states as the STATUS1 coefficient, and the difference between the executing states and the no death penalty states for the STATUS2 coefficient. As the output from REGRESSION in Figure 4 shows, this in indeed the case:

**Figure 4**

**---------------------------------------------------------------------------**

**Multiple R .49443**

**R Square .24446**

**Adjusted R Square .21231**

**Standard Error 3.46979**

**Analysis of Variance**

**DF Sum of Squares Mean Square**

**Regression 2 183.08974 91.54487**

**Residual 47 565.85446 12.03946**

**F = 7.60374 Signif F = .0014**

**------------------ Variables in the Equation ------------------**

**Variable B SE B Beta T Sig T**

**STATUS1 2.007143 1.135756 .257430 1.767 .0837**

**STATUS2 5.991071 1.537821 .567498 3.896 .0003**

**(Constant) 4.971429 .927341 5.361 .0000**

**---------------------------------------------------------------------------**

As noted earlier, most of the identities between various numbers on the output disappear when multiple predictors are involved. The t-tests for the individual STATUS coefficients indicate that the hypothetical population means for no death penalty and death penalty but no executions groups may not be different, while the hypothetical population mean for the executing states would seem to be higher than that for the no death penalty states. The F-test in the analysis of variance table tests the null hypothesis that both STATUS coefficients are 0 in the population, which means logically that all three population means are equal. Note that despite the logical relationships between hypotheses tested by the two types of tests (one or more nonzero coefficients implies rejection of the omnibus null hypothesis, and vice versa), when making inferences about population values based on sample data, it is possible to obtain logical contradictions. That is, one or more individual t-tests may be "significant" when the overall F-test is not, or the overall F-test may be "significant" when no individual t-tests are. If the overall F-test is "significant" then a paramterization can be found that will produce at least one "significant" t-test, but in some cases this parameterization many not provide any useful interpretation in terms of differences among group means.

Figure 5 shows what the default MANOVA parameterization produces for this analysis. As can be seen in the ANOVA table, the overall F-value for STATUS89 is identical to that given for the overall regression in REGRESSION. This is true because even though the different parameterizations produce different coefficients, we are still using two predictors to differentiate among three groups. Any two nonredundant predictors accurately representing the differences among our three groups would produce the same results. This is formally stated by saying that the overall test is invariant under different parameterizations. That is, we are testing the same omnibus null hypothesis of equality among the three population means, regardless of the specific contrast codings we use to compare groups. As with the earlier dichotomous predictor, the CONSTANT coefficient in this analysis represents the simple unweighted mean of the sample group means.

**Figure 5**

**---------------------------------------------------------------------------**

**Tests of Significance for MURDER90 using UNIQUE sums of squares**

**Source of Variation SS DF MS F Sig of F**

**WITHIN+RESIDUAL 565.85 47 12.04**

**STATUS89 183.09 2 91.54 7.60 .001**

**- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -**

**Estimates for MURDER90**

**--- Individual univariate .9500 confidence intervals**

**CONSTANT**

**Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper**

**1 7.63750000 .55726 13.70539 .00000 6.51643 8.75857**

**STATUS89**

**Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper**

**2 -2.6660714 .77278 -3.44996 .00119 -4.22071 -1.11143**

**3 -.65892857 .67370 -.97808 .33304 -2.01423 .69638**

**---------------------------------------------------------------------------**

The two deviation coefficients for STATUS89 give the means of each of the first two levels minus the unweighted grand mean of all three levels. This type of parameterization is often referred to as "effect" coding, because the parameter for a level of a factor variable is interpreted as the effect of being at that level of the predictor as opposed to being at the overall average. This terminology is probably more familiar to users whose linear models work has been handled in an analysis of variance framework, as opposed to a multiple regression approach. What is important to note here is that notwithstanding certain terminological preferences, both approaches are doing exactly the same thing.

Since we're fitting the same overall model in MANOVA that we did in REGRESSION, we must be able to predict the same values for each state that were predicted earlier. In order to be able to see exactly how this is done, we need to look at the design or basis matrix that MANOVA used. As you can see by looking at the coefficients in matrix in Figure 6, the predicted value for the no death penalty group is derived by summing the first and second parameter estimates, the value for the death penalty but no executions group is obtained by summing the first and third estimates, and the death penalty group value is obtained by subtracting the second and third estimates from the first. By doing the arithmetic here yourself, you can verify the identical predictions resulting from the two different parameterizations.

**Figure 6**

**---------------------------------------------------------------------------**

**STATUS89 CONSTANT STATUS89 STATUS90**

**(1) (2) (3)**

**0 1 1 0**

**1 1 0 1**

**2 1 -1 -1**

**---------------------------------------------------------------------------**

**INTERACTIONS AMONG DICHOTOMOUS PREDICTORS IN REGRESSION   
David P. Nichols   
Senior Support Statistician   
SPSS, Inc.   
August 1995**

Here we continue the discussion of parameterization of linear regression models involving categorical predictor variables from Keywords issues 56 and 57. In this article we will deal with the problem of interactions among categorical predictors. We will continue to use the United States data with 1990 murder rate as our dependent variable, and also to suppose that the states might be viewed as a random sample from some theoretical population of interest, in order to motivate attention to test statistics.

We'll deal with the simplest case of interaction, that between two binary or dichotomous predictor variables. DEATHPEN, introduced in issue 56, is a 0-1 variable indicating absence or presence of a death penalty statute in 1989-90. CULTURE is a new variable, again 0-1, indicating absence or presence of a certain level of influence of a set of cultural characteristics thought by some social scientists to be involved in the production of high rates of certain types of violence. Though both predictor variables are measured prior to the dependent variable and are potentially causal contributants, there are undoubtedly a number of other factors left out of our simple model. Thus, the relationships shown here should be taken only as an illustration of basic regression methods and not as a rigorous analysis of the contributants to murder rates.

Figure 7 contains the means, standard deviations and cell counts for the four combinations of the two predictors. We can see that the means for states without the cultural factor are lower than those with it, and that on average the states without the death penalty have lower means. However, we also see that the mean for the CULTURE 0 states is higher for the states without the death penalty. There is thus evidence of a differential impact of DEATHPEN depending on what level of CULTURE one considers. In other words, it would appear that DEATHPEN interacts with CULTURE in it's effects on MURDER90. Formally, an interaction means that the effect of a predictor on the dependent variable is dependent upon the level of the other predictor considered. (Incidentally, the 13 states in the CULTURE 0-DEATHPEN 1 cell do not include any of the states carrying out executions in 1989. Though a 2x3 analysis using the three level STATUS89 variable from issue 57 is perhaps a more theoretically satisfactory one, it results in an empty combination of predictors, which produces issues too complicated to be dealt with in this brief article.)

**Figure 7: Descriptive Statistics**

**---------------------------------------------------------------------------**

**Variable .. MURDER90**

**FACTOR CODE Mean Std. Dev. N**

**CULTURE 0**

**DEATHPEN 0 4.840 4.307 10**

**DEATHPEN 1 4.092 1.506 13**

**CULTURE 1**

**DEATHPEN 0 5.300 1.671 4**

**DEATHPEN 1 9.996 2.796 23**

**---------------------------------------------------------------------------**

We will first analyze the data using the REGRESSION procedure, entering the two dummy variables CULTURE and DEATHPEN, and a product variable computed by multiplying the two variables. This INTERACT product variable, when entered along with CULTURE and DEATHPEN, represents the interaction of CULTURE and DEATHPEN. The results of the regression are given in Figure 8.

**Figure 8: REGRESSION results with dummy coding**

**---------------------------------------------------------------------------**

**Multiple R .70702**

**R Square .49988**

**Adjusted R Square .46726**

**Standard Error 2.85354**

**Analysis of Variance**

**DF Sum of Squares Mean Square**

**Regression 3 374.38140 124.79380**

**Residual 46 374.56280 8.14267**

**F = 15.32591 Signif F = .0000**

**------------------ Variables in the Equation ------------------**

**Variable B SE B Beta T Sig T**

**CULTURE .460000 1.688175 .059237 .272 .7865**

**DEATHPEN -.747692 1.200261 -.086742 -.623 .5364**

**INTERACT 5.443344 1.957121 .700974 2.781 .0078**

**(Constant) 4.840000 .902367 5.364 .0000**

**---------------------------------------------------------------------------**

The most important thing to notice here is that the INTERACT variable has a significance level of .0078, indicating that the interaction term should remain in the model. Some people might look at the significance levels for the CULTURE and DEATHPEN "main effects," which are both well above .05, and conclude that we have a situation where this is an interaction but no main effects. This evinces a misunderstanding of the meaning of an interaction. An interaction means that the effects of a variable differ across the levels of another variable. In order for the effects of one variable to differ at different levels of another variable, some of these effects must be nonzero. Logically then, an interaction implies that all involved main effects are present as well.

An important feature of the interaction model is brought out by comparing these results with those from the MANOVA procedure, where we've fitted the same model, but used a somewhat different parameterization. Recall that in our previous analyses involving only main effects, the parameterization did not change the overall main effects F-test (only the constant was affected). We see here that this no longer holds once an interaction is introduced into the model. Figure 9 presents the (edited) MANOVA results using SIMPLE(1) contrasts, which request comparison of the second category of each factor to the first, just as is produced by dummy coding when we fit only main effects.

**Figure 9: MANOVA results with SIMPLE(1) contrasts**

**---------------------------------------------------------------------------**

**Parameter Coeff. Std. Err. t-Value Sig. t Lower -95% CL- Upper**

**CONSTANT 6.05698997 .48928 12.37939 .00000 5.07212 7.04186**

**CULTURE 3.18167224 .97856 3.25138 .00215 1.21193 5.15141**

**DEATHPEN 1.97397993 .97856 2.01723 .04953 .00424 3.94372**

**CULTURE BY 5.44334448 1.95712 2.78130 .00782 1.50386 9.38282**

**DEATHPEN**

**---------------------------------------------------------------------------**

Again, the important thing to look at is the interaction term, which is identical to that given by REGRESSION. Note that changing the parameterization might change the scaling of the parameter here, but would not change the value of the t-statistic or it's significance. Relationships among individual parameters are more complicated when factors have more than two levels, but the overall F-statistics remain the same regardless of parameterization when the highest order term is being considered. Since all terms here have one degree of freedom, the F-tests test the same thing as the t-tests, and have been omitted to save space.

The next thing we might notice is that according to the MANOVA results, both of the "main effects" are significant, unlike in the REGRESSION results. This is where understanding how the model has been parameterized is crucial. Let's look at the "parameter codings" or basis matrices used in the two analyses to see how we can square the two sets of findings. Figure 10 gives the values of the codings for the dummy approach. The 4x4 matrix on the right is the basis or design matrix (X) used in the linear model. The contrast matrix (C) produced by this coding scheme is given in Figure 11. The relationship between the two matrices can be verified by evaluating the following equation:

**C = (X'X)-1X'**

**Figure 10: Parameterization using dummy codings**

**---------------------------------------------------------------------------**

**CULTURE DEATHPEN | CONSTANT CULTURE DEATHPEN INTERACT**

**0 0 | 1 0 0 0**

**0 1 | 1 0 1 0**

**1 0 | 1 1 0 0**

**1 1 | 1 1 1 1**

**---------------------------------------------------------------------------**

Reading across the first row of Figure 10, we can see that the CULTURE 0, DEATHPEN 0 states are represented completely by the CONSTANT parameter. This is borne out by the values in the first row of the contrast matrix in Figure 11, where the CONSTANT is seen to be simply the mean of the CULTURE 0, DEATHPEN 0 group, and by noting that the 4.84 value for the CONSTANT in the REGRESSION output is the mean for those states. The second row of the basis shows that the CULTURE 0, DEATHPEN 1 mean is modeled by summing the CONSTANT and DEATHPEN parameters, which implies that the DEATHPEN parameter compares this group to the CULTURE 0, DEATHPEN 0 group. The third line of the contrast matrix shows that the DEATHPEN parameter is indeed comparing these two groups. Again, you should be able to derive the parameter estimate value from the appropriate means (within printed levels of precision). Thus the DEATHPEN effect here is really the simple main effect of DEATHPEN at the 0 level of CULTURE (4.092-4.84=-.748), which according to the significance level printed is quite possibly chance variation. The third row of the basis in Figure 10 and the second row of the contrast matrix in Figure 11 show that the CULTURE parameter is assessing the simple main effect of culture for the DEATHPEN 0 states (5.3-4.84=.46), which is also easily attributable to chance. Finally, the INTERACT parameter estimates the difference between the simple main effects of CULTURE at the two levels of DEATHPEN and vice versa:

5.44 = (9.996-4.092) - (5.3-4.84) = (9.996-5.3) - (4.092-4.84).

Thinking about the interaction parameter in this way illustrates how an interaction implies main effects: interactions are differences among differences, and if all differences are 0, then by definition the differences among the differences must also be 0.

**Figure 11: Contrasts estimated by dummy codings**

**---------------------------------------------------------------------------**

**CULTURE0 CULTURE0 CULTURE1 CULTURE1**

**DEATHPEN0 DEATHPEN1 DEATHPEN0 DEATHPEN1**

**CONSTANT 1 0 0 0**

**CULTURE -1 0 1 0**

**DEATHPEN -1 1 0 0**

**INTERACT 1 -1 -1 1**

**---------------------------------------------------------------------------**

Now let's look at the basis and contrast matrices for the SIMPLE(1) coding used in MANOVA (Figure 12). The last line of Figure 13 shows that the contrast estimated by the interaction parameter is the same as with dummy coding, which fits with our earlier observation that the parameter estimates for the interaction terms were the same in both analyses. As we noted earlier though, none of the other parameters represent the same things. The similarity of the basis and contrast matrices for the SIMPLE contrasts is an artifact of the 2x2 design, and does not hold for designs involving larger numbers of levels. In general, it is difficult to use the basis matrix with SIMPLE contrasts to see what is being estimated.

**Figure 12: Parameterization using SIMPLE(1) codings**

**---------------------------------------------------------------------------**

**CULTURE DEATHPEN | CONSTANT CULTURE DEATHPEN INTERACT**

**0 0 | 1 -.5 -.5 .25**

**0 1 | 1 -.5 .5 -.25**

**1 0 | 1 .5 -.5 -.25**

**1 1 | 1 .5 .5 .25**

**---------------------------------------------------------------------------**

The contrast matrix shows us that the CONSTANT parameter is simply estimating the (unweighted) average of the four means. The CULTURE parameter estimates the average of the two CULTURE 1 cells minus the average of the two CULTURE 0 cells, while the DEATHPEN parameter estimates the average of the two DEATHPEN 1 cells minus the average of the two DEATHPEN 0 cells. You should be able to use the cell means to reproduce the parameter estimate values. Note that what are labeled as main effects here are averages of the simple main effects of each factor across the levels of the other factor. Thus the 1.974 coefficient for DEATHPEN is the average of the 9.996-5.3=4.696 difference at level 1 of CULTURE and the 4.092-4.84=-.748 difference at level 0 of CULTURE. Though such "main effects" have become the norm for computer output from ANOVA/linear models procedures, largely due to the simplicity of the hypotheses tested, the danger in taking averages of different effects as representative of the whole is well illustrated by this example. If DEATHPEN were a treatment, CULTURE denoted two types of patients and the dependent variable were a measure of health, for example, we might conclude based on the averaged effect that the treatment is good for everyone, when the results are really telling us that it has no effect or is perhaps harmful for one group of patients.

**Figure 13: Contrasts estimated by SIMPLE(1) codings**

**---------------------------------------------------------------------------**

**CULTURE0 CULTURE0 CULTURE1 CULTURE1**

**DEATHPEN0 DEATHPEN1 DEATHPEN0 DEATHPEN1**

**CONSTANT .25 .25 .25 .25**

**CULTURE -.50 -.50 .50 .50**

**DEATHPEN -.50 .50 -.50 .50**

**INTERACT 1.00 -1.00 -1.00 1.00**

**---------------------------------------------------------------------------**

In addition to using the means to verify the interpretations of the parameter estimates, you should be able to use the basis matrices and parameter estimate values to reproduce the predicted values for each cell for both analyses. If you do this, you will see that the predictions produced in each case are identical. This is true because have fitted exactly the same model in both cases. The choice of parameterization strategy affects only the interpretation of the individual parameters, not the overall model. The most important thing to note here is that the interaction is the only term that is independent of the choice of parameterizations. The presence of an interaction means that there is no single main effect for a factor involved in that interaction, so different choices of parameterization lead to different interpretations for the "main effects."

**CONTINUOUS BY CATEGORICAL INTERACTIONS IN REGRESSION   
David P. Nichols   
Senior Support Statistician   
SPSS, Inc.   
From SPSS Keywords, Number 61, 1996**

Continuing the topic of using categorical variables in linear regression, in this issue we will briefly demonstrate some of the issues involved in modeling interactions between categorical and continuous predictors. As in previous issues, we will be modeling 1990 murder rates in the 50 states of the U. S. Our predictors will be the previously used 0-1 culture dummy variable, along with a new variable: state 1990 per capita income, expressed as a percentage deviation from the national average (i.e., a value of 10 indicates that a state's per capita income was 10% above the national mean).

For most people, the parameterization of choice in this situation is to code the culture dummy variable so that 0 means no and 1 means yes (that state is deemed to be affected by the cultural factor of interest). The interaction variable is created by multiplying the dummy variable by the income variable. Results of this REGRESSION are given in Figure 14. As always, the constant is the predicted value for the dependent variable when all predictors are 0. In this case, it represents the predicted 1990 murder rate for a state without the cultural characteristic of interest, with 1990 per capita income equal to the national average. The culture parameter gives the change in predicted value for the affected states relative to the others when income is 0 (at the national average). Affected states have a much larger predicted rate (almost twice as high). The income parameter gives the predicted slope for the unaffected states; increases in per capita income are associated with higher predicted murder rates for these states. The interaction parameter estimates the difference in predicted slope for the affected states; the -.13 value here can be added to the .08 value for income to obtain the predicted slope for these states, which is about -.047. In other words, for these states, increases in income are associated with decreases in predicted murder rates.

**Figure 14: REGRESSION results with dummy coding (1=YES)**

**----------------------------------------------------------------------------**

**Variable B SE B Beta T Sig T**

**CONSTANT 4.530989 .620501 7.302 .0000**

**CULTURE 4.377789 .909829 .563756 4.812 .0000**

**INCOME .082682 .040147 .318512 2.059 .0451**

**CUL\_INC -.130178 .057808 -.366026 -2.252 .0291**

**----------------------------------------------------------------------------**

Note that the interpretation of the "main effects" is conditional upon the level of the other variable, due to the presence of the interaction in the model. An alternative parameterization is presented in Figure 15. Here, the dummy variable for culture has been reverse coded, 0 for affected, 1 for not. Here, the constant now gives the affected group's predicted value at 0 or mean income. Culture again compares the two groups at mean income, but this time subtracts affected from unaffected and thus has a negative sign. Income gives the -.047 value we calculated above: the predicted slope for the affected states. The interaction coefficient is the same but with an opposite sign. Again, adding this to income produces the predicted slope for the group coded 1.

**Figure 15: REGRESSION results with dummy coding (1=NO)**

**----------------------------------------------------------------------------**

**Variable B SE B Beta T Sig T**

**CONSTANT 8.908778 .665407 13.388 .0000**

**CULTURE -4.377789 .909829 -.563756 -4.812 .0000**

**INCOME -.047495 .041593 -.182964 -1.142 .2594**

**CUL\_INC .130178 .057808 .351942 2.252 .0291**

**----------------------------------------------------------------------------**

Note that the t-statistics for the constant and income parameters differ between the two tables. This is because these parameters are estimates of different things under the alternative parameterizations. You should be able to reproduce exactly the same predicted value for a given combination of culture and income from the two parameterizations in order to verify the fact that they are estimating the same overall model. For example, for an affected state with an income value of 10, parameterization 1 would predict

4.530989 + 4.377789 \* 1 + .082682 \* 10 - .130178 \* 1 \* 10 = 8.4338

while parameterization 2 would predict

8.908778 - 4.377789 \* 0 - .047495 \* 10 + .130178 \* 0 \* 10 = 8.4338.

While the intercept and income terms here varied with the parameterization of the model, the culture term did not. There are ways of parameterizing the model that would produce different results for culture. Can you think of what some of these might be? We'll talk about this in the next issue.

**FURTHER INTERACTIONS WITH CATEGORICAL VARIABLES IN REGRESSION  
David P. Nichols  
Senior Support Statistician  
SPSS, Inc.  
From SPSS Keywords, Number 62, 1996**

In the last issue, we talked about interpretation of the parameters of a regression model that included one dichotomous and one continuous predictor, plus their interaction. As has been the case throughout this series, the point was to illustrate the dependence of parameter interpretations on the way the predictor variables were coded; that is, on how the model was parameterized. We saw how the interpretation of the "main effect" of the continuous income variable was conditional upon the way the dichotomous culture variable was coded. We also saw that in the two parameterizations we used (reversing the 0-1 dummy coding for culture), the culture "main effect" parameter had the same absolute value, producing the same t-statistic and significance level. At the end of that article, we alluded to the fact that this would not necessarily be the case under alternative parameterizations, and promised to show why.

Figure 16 gives the same numbers as Figure 14 (from the above article). Recall that the culture dummy variable was coded 0 for no and 1 for yes, and that the income variable was expressed as a percentage deviation from the national mean. Though it was not specified at the time, the national mean was based on an unweighted mean of individuals or a weighted mean of states, so that states with higher populations contributed more heavily. When we use such a variable in an analysis with each state treated as a single unit, the mean of the income variable is not 0; in this case, it's -5.08 (states with larger populations tend to have higher relative incomes).

**Figure 16: Original REGRESSION results (from previous issue)**

**-------------------------------------------------------------------------------**

**Variable B SE B Beta T Sig T**

**CONSTANT 4.530989 .620501 7.302 .0000**

**CULTURE 4.377789 .909829 .563756 4.812 .0000**

**INCOME .082682 .040147 .318512 2.059 .0451**

**CUL\_INC -.130178 .057808 -.366026 -2.252 .0291**

**-------------------------------------------------------------------------------**

Suppose we now recompute the income variable by centering it; that is, we make it so that it has a mean of 0 in our sample (we do this by adding 5.08 to the earlier income variable). What happens now when we recompute the interaction product variable for culture and income and run a regression using the same dummy coding for culture? As you can see from Figure 17, we now get different results for the constant term and for the culture "main effect." The constant is of course the predicted value when all predictors are set to 0. In both cases, this means culture=0. However, in Figure 16, it means that income is set to 100% of the national mean, while in Figure 17, income is set to 94.92% of the national mean. Thus the constant for Figure 17 is equal to the original constant minus 5.08 times the income coefficient:

4.110964 = 4.530989 - 5.08 \* .082682.

**Figure 17: Centered REGRESSION results**

**-------------------------------------------------------------------------------**

**Variable B SE B Beta T Sig T**

**CONSTANT 4.110964 .635702 6.467 .0000**

**CULTURE 5.039091 .864147 .648916 5.831 .0000**

**INCOME .082682 .040147 .318512 2.059 .0451**

**CUL\_INC -.130178 .057808 -.343087 -2.252 .0291**

**-------------------------------------------------------------------------------**

Notice also that the culture coefficient has changed, from 4.377789 to 5.039091, as has it's t-value and significance. This is because we are now estimating something different: the difference in predicted value for a state with culture=1 compared with a state with culture=0, but this time at income equal to -5.08 on the old scale (which is 0 on the new scale). The culture coefficient in Figure 17 is (to within rounding error) the original culture coefficient minus 5.08 times the interaction coefficient:

5.039091 = 4.377789 - 5.08 \* (-.130178).

The primary implication of course is that the interpretation of the "main effect" of the culture variable, like that of the income variable, depends on how the model has been parameterized. There is no single interpretation of this effect available. Another implication is that we can make this coefficient estimate the predicted difference between groups at any fixed value of the income variable we choose, simply by subtracting that value from the original variable. A common usage of this property is the centering of variables so that comparisons are made at the mean of a variable rather than at the 0 point of the original continuous predictor, which often isn't of interest.

Finally, you may have noticed that the interaction coefficient remained the same in absolute value throughout our variable transformations. It is possible to change this value by rescaling the predictors. If we change the distance between the two groups on the dichotomous culture predictor, while keeping the continuous income predictor the same, the result is to multiply the interaction coefficient by the reciprocal of the change in distance (e.g., doubling the distance between the two group codes, to say -1 and 1 rather than 0 and 1, results in a halving of the interaction coefficient). Similarly, keeping the unit distance between the culture codings and multiplying the continuous income predictor by a constant produces a reciprocal change in the interaction coefficient (e.g., multiplying the income variable by two produces an interaction coefficient half the size of the original one). Note in each case that the interaction product variable must be recomputed from the transformed original variables.

While linear transformations (multiplication by a constant and addition of another constant, or new=a+b\*old) of original variables will rescale the interaction term, the standard error will also be rescaled, resulting in the same t-statistic and significance level. The interaction term in this model is the highest order term in a hierarchical model, and is thus \_invariant\_ under such transformations. It is the only term in this model for which this is true. All lower order terms are "contained within" or "marginal to" this interaction effect, and are thus dependent upon the specific model parameterization for their meaning. In a nutshell, this is the lesson of this series.

**This page was adapted from a web page at the** [**SPSS web page**](http://www.spss.com)**.  We thank SPSS for their permission to adapt and distribute this page via our web site.**