INTERACTION EFFECTS IN MULTIPLE REGRESSION

SECOND EDITION

James Jaccard
Robert Turrisi

Series: Quantitative Applications in the Social Sciences

72

a SAGE UNIVERSITY PAPER

SAGE UNIVERSITY PAPERS

Series: Quantitative Applications in the Social Sciences

INSTRUCTIONS TO POTENTIAL CONTRIBUTORS

For guidelines on submission of a monograph proposal to this series, please write

Michael S. Lewis-Beck, Editor
Sage QASS Series
Department of Political Science
University of Iowa
Iowa City, IA 52242

Series/Number 07-72

INTERACTION EFFECTS IN MULTIPLE REGRESSION, Second Edition

JAMES JACCARD
University at Albany
State University of New York

ROBERT TURRISI
Boise State University

SAGE PUBLICATIONS
International Educational and Professional Publisher
Thousand Oaks London New Delhi

For information:

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

Sage Publications Ltd.
6 Bonhill Street
London EC2A 4PU
United Kingdom

Sage Publications India Pvt. Ltd.
B-42 Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Jaccard, James.
Interaction effects in multiple regression / James Jaccard, Robert Turrisi.—2nd ed.
p. cm.—(Quantitative applications in the social sciences; 72)
Includes bibliographical references and index.
ISBN 0-7619-2742-5 (pbk.)
1. Regression analysis. 2. Social sciences-Statistical methods.
I. Turrisi, Robert. II. Title. III. Series: Sage university papers series.
Quantitative applications in the social sciences; no. 07-072.
HA31.3 .J33 2003
519.5′36′0243—dc211

2002153223

06 10 9 8 7 6 5 4 3 2

Editorial Assistant:	Karen Wiley
Production Editor:	Melanie Birdsall
Copy Editor:	Connie Adams
Typesetter:	C&M Digitals (P) Ltd.
Proofreader:	Cheryl Rivard

When citing a university paper, please use the proper form. Remember to cite the Sage University Paper series title and include paper number. One of the following formats can be adapted (depending on the style manual used):

(1) JACCARD, J., and TURRISI, R. (2003). *Interaction Effects in Multiple Regression (2nd ed).* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-072. Thousand Oaks, CA: Sage.

OR

(2) Jaccard, J., and Turrisi, R. (2003). *Interaction effects in multiple regression, (2nd ed).* (Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-072). Thousand Oaks, CA: Sage.

CONTENTS

SERIES EDITOR'S INTRODUCTION

The classical regression model is straightforward. Take the usual three-variable, multiple regression example, with Y the dependent variable, X_1 and X_2 the independent variables. We may write Equation 1 as

$$Y = a + b_1X_1 + b_2X_2 + e \qquad [1]$$

where a, b_1, and b_2 are population parameters to be estimated and e represents the error term. Assuming the necessary assumptions are met, then the ordinary least squares (OLS) estimator would be the best linear unbiased estimator (BLUE). This OLS equation appears additive in that the terms are added up to account for Y and none of the variables themselves are multiplied together. But suppose the independent variables are multiplied together, forming a product term, X_1X_2, and the equation is rewritten as Equation 2,

$$Y = a + b_1X_1 + b_2X_2 + b_3(X_1X_2) + e \qquad [2]$$

We now have an equation for which variable values are additive and nonadditive. The product, or multiplicative, term (X_1X_2) is also called an interaction term. Its coefficient, b_3, estimates the interaction effect. The question is, what does it mean?

Usually, first-year graduate students in the social sciences find this question hard to answer. They understand the interpretation of effects in the additive model of Equation 1: b_1 indicates the expected change in Y for a unit change in X_1, holding the X_2 constant. But how do we interpret b_3, the coefficient of the nonadditive term in Equation 2? When the effect of X_1 on Y depends on the value of X_2, an interaction effect is present. That interaction effect is estimated by b_3. Consider this illustration: Y = individual contributions to political campaigns measured in dollars, X_1 = income in dollars, X_2 = education (scored 0 = no college, 1 = at least some college). If the researcher makes the plausible argument that the effect of income on contributions is greater among the college educated, then the interaction specification of Equation 2 is preferred to Equation 1.

Obviously, there are many times when the interaction hypothesis—the *impact* of one variable depends on the *value* of another—would merit testing. But in a review of published studies, it does not pop up that often. That, I suggest, is because the interaction idea can be hard to grasp and hence the singular importance of this second edition. Dr. Jaccard and Dr. Turrisi make the interpretation and estimation of interaction effects in a regression model crystal clear. As a glance at the table of contents shows, this volume is significantly changed and updated from the first edition, which appeared in 1990. The valuable old points remain. For instance, they emphasize that it is simply inadequate to examine interactions by looking within the separate categories of the moderator variable, for example, in our illustration above, looking at a regression of campaign contributions on income within the college and noncollege groups. Further, they provide an often forgotten insight: The true metric is in the data, not the measure. Hence, for testing interactions with regression, what is important is the extent to which the measure comes close to an underlying interval property in the data. Also, there are many new topics, such as interaction models with clustered data and random coefficient models. Indeed, perusing the bibliography, one can count 30 references to works appearing since the first edition.

The practicing research worker must be able to posit, interpret, and estimate interaction effects with facility. Careful study of this volume enables that to happen in the regression context. With regard to the study of interactions in other, perhaps more advanced, research contexts, the reader should consult the other interaction monographs Professor Jaccard has authored or coauthored in this series, namely, "LISREL Approaches to Interaction Effects in Multiple Regression," No. 114, and "Interaction Effects in Factorial Analysis of Variance," No. 118.

—*Michael S. Lewis-Beck*
Series Editor

PREFACE

This revised edition of *Interaction Effects in Multiple Regression* has the same intent as the first edition, namely, to introduce the reader to the basics of interaction analysis using multiple regression methods with one or more continuous predictor variables. The monograph is neither a technical nor an advanced exposition of this complex topic. Our goal is to present a nontechnical, introductory orientation to the interpretation of traditional interaction models that use product terms. The monograph is oriented toward the researcher with rudimentary background in multiple regression. We have avoided complex formulas, which can be intimidating to applied researchers. As an alternative, we have provided the reader with simple (but sometimes cumbersome) computer-based heuristics that permit the calculation of parameter estimates and estimated standard errors that typically will be of interest. The first three chapters are elementary, with Chapter 4 touching on a wide range of more advanced issues. The intent of Chapter 4 is to alert readers to relevant advanced issues, give them an overview of the advanced issues, and then provide them with references that permit them to follow up on the issues in more detail.

We would like to thank the anonymous reviewers who provided us with feedback on this revision as well as the series editor, Michael Lewis-Beck, for his perceptive and useful commentary and support.

INTERACTION EFFECTS IN MULTIPLE REGRESSION, SECOND EDITION

JAMES JACCARD
University at Albany
State University of New York

ROBERT TURRISI
Boise State University

1. INTRODUCTION

Many theoretical frameworks in the social sciences focus on causal models. These models specify the effects of one or more independent variables on one or more dependent variables or outcome variables. At the simplest level, there are six types of relationships that can occur within a causal model, as illustrated in Figure 1.1. A *direct* causal relationship is one in which a variable, *X*, is a direct cause of another variable, *Y*. It is the immediate determinant of *Y* within the context of the theoretical system. An *indirect* causal relationship is one in which *X* exerts a causal impact on *Y* but only through its impact on a third variable, *Z*. A *spurious* relationship is one in which *X* and *Y* are related but only because of a common cause, *Z*. There is no formal causal link between *X* and *Y*. A *bidirectional* or *reciprocal* causal relationship is one in which *X* has a causal influence on *Y*, which in turn has a causal impact on *X*. An *unanalyzed* relationship is one in which *X* and *Y* are related but the source of the relationship is unspecified. Finally, a *moderated* causal relationship is one in which the relationship between *X* and *Y* is moderated by a third variable, *Z*. In other words, the nature of the relationship between *X* and *Y* varies, depending on the value of *Z*.

This monograph is concerned with the statistical analysis of moderated relationships and focuses on the case where one or more of the independent variables (or predictor variables) is continuous in nature. Moderated relationships often are called interaction effects, although precise conceptualizations of interaction effects vary across statistical models. Our focus is on analyzing moderated relationships in multiple regression. There currently exists confusion about the analysis of moderated relationships involving continuous

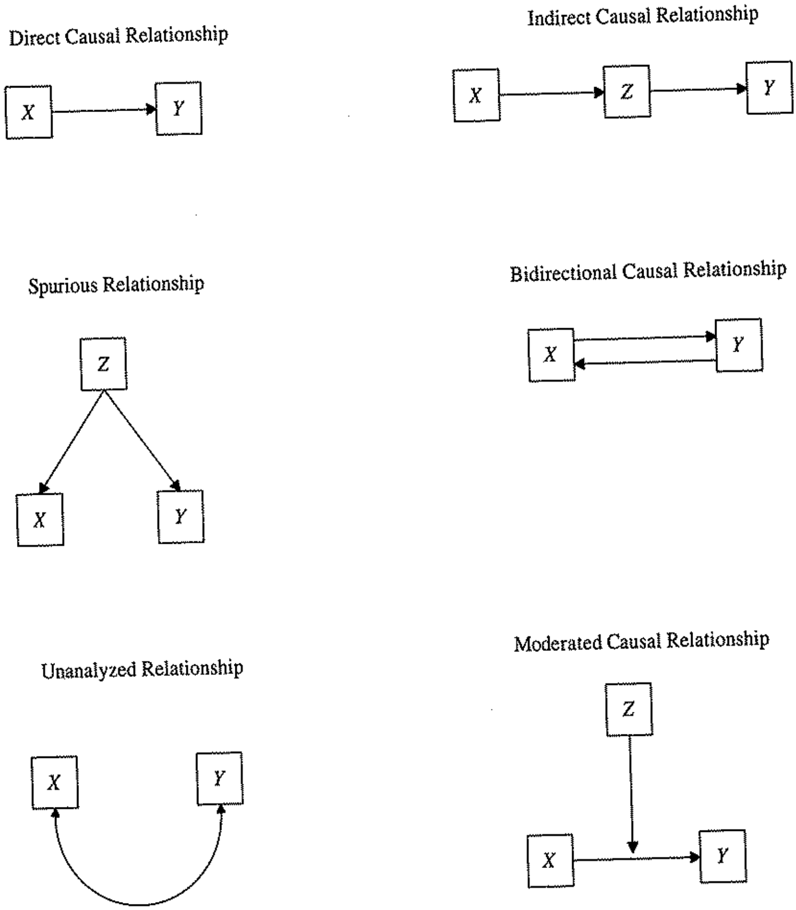

Figure 1.1 Examples of Causal Relationships

variables using multiple regression-based methods. The substantive literature is replete with contradictory advice and admonitions about the best way to test such models. The major purpose of this monograph is to bring together the rather diverse literature on this topic and to explicate the central issues involved in conducting analyses of moderated relationships involving continuous variables. Our goal is to present a readable and practical introduction for the social science researcher who has working knowledge of multiple regression.

In this chapter, we briefly review key concepts in the analysis of interactions in social science research. We begin by considering the concept of interaction in general and then explicate the concept of a simple main effect and an interaction contrast. We then review basic multiple regression procedures in order to provide a framework for future chapters, including the use of dummy variables, the effects of simple transformations on regression coefficients, and the calculation of predicted scores.

The Concept of Interaction

As noted, there are many ways in which interaction effects have been conceptualized in the social sciences and there is controversy about the best way to think about the concept. One popular school of thought conceptualizes interaction effects in terms of moderated relationships. This perspective can be illustrated using a three-variable system in which one of the variables is construed as an outcome variable, a second variable is viewed as an independent variable, and a third variable is viewed as a moderator variable. In this system, the outcome variable is thought to be influenced by the independent variable. An interaction effect is said to exist when the effect of the independent variable on the dependent variable differs depending on the value of a third variable, called the moderator variable. For example, the effect of the number of years of education on income may differ depending on one's ethnicity. Education may have a larger impact on income for some ethnic groups than for other ethnic groups. In the moderator framework, income is the outcome variable, the number of years of education is the independent variable, and ethnicity is the moderator variable. As another example, the effect of social class on how often someone uses a health clinic may vary depending on gender. In this case, how often someone uses a health clinic is the outcome or dependent variable, social class is the independent variable, and gender is the moderator variable. Gender is said to "moderate" the effect of social class on clinic use.

The moderator approach to interaction analysis requires that a theorist specify a moderator variable and what we call a focal independent variable. The focal independent variable is the variable whose effect on the dependent variable is thought to vary as a function of the moderator variable. Most formal research questions readily lend themselves to the specification of one of the predictors as having "moderator" status. For example, a researcher might want to determine if a clinical treatment for depression is more effective for males than for females. It is evident in this case that gender is the moderator variable and the presence versus absence of the treatment is the focal independent variable.

Situations arise where one theorist's moderator variable is another theorist's focal independent variable and vice versa. For example, a consumer psychologist who studies product quality and product choice might be interested in the effect of product quality on product purchase decisions and how this is moderated by the pricing of products. In contrast, a marketing researcher using the same experimental paradigm as the consumer psychologist might be interested in the effect of product pricing on product purchase decisions and how this is moderated by product quality. In both cases, the designation of the moderator variable follows from the practical and theoretical orientations of the researchers. Neither specification is better than the other, and, as we will see, statistically, the results of an interaction analysis will be the same in the two conceptualizations. The two designations simply represent different perspectives on the same phenomena and guide the researchers to emphasize different aspects of the data.

Situations also arise where the theorist is unsure which variable should have moderator status. Suppose a researcher is examining the effects of gender and ethnicity on attitudes toward abortion. He or she might want to characterize how gender differences in such attitudes vary as a function of ethnicity. In this case, gender is the focal independent variable and ethnicity is the moderator variable. The researcher also might be interested in characterizing ethnic differences in attitudes toward abortion and how these ethnic differences vary as a function of gender. In this case, ethnicity is the focal independent variable and gender is the moderator variable. There is nothing to prevent the researcher from characterizing the data from both perspectives in such scenarios. It is all a matter of what is of substantive interest. We return to this issue in more depth in Chapter 4.

Other approaches for conceptualizing statistical interaction are discussed in Jaccard (1998) and Jaccard and Dodge (2003). The moderator approach is commonly invoked in substantive research domains, even though researchers sometimes do not even realize they are doing so. Interaction effects can be difficult to imbue with substantive meaning in an applied research setting, and most researchers who do so successfully ultimately fall back on the moderator framework. Given its popularity and conceptual ease, we adopt the framework in this monograph. In Chapter 4, we revisit alternative frameworks for thinking about interactions.

Simple Effects and Interaction Contrasts

Two important concepts in moderator analysis are *simple effects* (also called *simple main effects*) and *single degree-of-freedom interaction contrasts*. These concepts are best illustrated using qualitative predictor variables in a factorial design. Suppose a social scientist identifies 300 married individuals, half of whom are male and half of whom are female (but they are not

TABLE 1.1
Attitudes Toward Abortion as a Function of Gender and Religion

	Catholic	*Protestant*	*Jewish*
Females	5.0	6.0	7.0
Males	3.0	3.0	3.0

married to one another). One third of the individuals are Catholic, one third are Protestant, and one third are Jewish. This yields a 2×3 factorial design that crosses gender by religion. The researcher is interested in how gender and religion are related to attitudes toward abortion, which is measured on a 0 to 10 scale, with higher scores indicating more favorable attitudes. The mean values for each subgroup are presented in Table 1.1. Suppose that, in this particular case, the researcher decides to conceptualize religion as the moderator variable and gender as the focal independent variable.

Simple Effects

There are many questions that the researcher can pose of these data. One common question is whether the focal independent variable (gender) has an effect on the outcome variable at each level of the moderator variable considered separately. Answering this question involves comparing the mean for males with the mean for females just for Catholics, doing so again just for Protestants, and then doing so again just for Jews. That is, the researcher conducts three significance tests, one at each level of the moderator variable. These contrasts are commonly called simple effects or simple main effects. They focus on the effect of the focal independent variable on the outcome variable at a given level of the moderator variable. For the data in Table 1.1, the simple effect of gender for Catholics is a test of significance of the mean difference of $5.0 - 3.0 = 2.0$.

Another way of thinking about simple effects is that they are conditional effects. A simple effect is the effect of the independent variable on the outcome variable conditioned on the moderator variable being equal to a particular value (e.g., conditioned on the moderator variable being equal to the value "Jewish").

Interaction Contrasts

Simple main effects often are of conceptual interest, but they have little to do with interaction effects (even though some researchers think they do—a point to which we return later). For an interaction effect to exist in the moderator framework, the effect of the focal independent variable on the outcome variable must differ depending on the level of the moderator variable. For example, the gender difference between males and females must be different for Catholics than it is for Protestants or different for Catholics than it is for Jews or different for Protestants than it is for Jews. For the data in Table 1.1, the gender difference for Catholics is the female mean minus the male mean focusing on just the Catholics, or 5.0 – 3.0 = 2.0. The gender difference for Protestants is the female mean minus the male mean focusing on just the Protestants, or 6.0 – 3.0 = 3.0. The gender difference for Catholics (2.0) is smaller than the gender difference for Protestants (3.0), and this suggests an interaction effect; that is, the effect of gender on attitudes toward abortion depends on the religion of the individual. This effect can be captured in a single number by calculating the difference between the two mean differences, 2.0 – 3.0 = 1.0. The fact that this estimated interaction parameter is not zero suggests that an interaction effect is present. Of course, the nonzero value may simply reflect sampling error, so a formal significance test of the estimated parameter would need to be performed.

The above example illustrates what is called a single degree-of-freedom interaction contrast. It is an interaction contrast because it explicitly compares the effect of the focal independent variable on the outcome variable at one level of the moderator variable with the corresponding effect at another level of the moderator variable. The gender difference for Catholics (2.0) was formally contrasted with the gender difference for Protestants (3.0). Because the statistical test of the contrast has only a single degree of freedom in the numerator, it is called a single degree-of-freedom interaction contrast. For greater discussion of such contrasts, see Jaccard (1996).

Single degree-of-freedom interaction contrasts are at the heart of interaction analysis. They represent focused tests of the interaction. They are distinguished from omnibus interaction tests, which can involve more than a single degree of freedom. Omnibus interaction significance tests are global tests of interaction that focus on the independent variable and the moderator variable in their entirety rather than on subgroups within them. In the factorial design of Table 1.1, the omnibus test of interaction focuses on the overall interaction between gender and religion, which in this case has 2 degrees of freedom (*df*). By contrast, the example we used for a single degree-of-freedom interaction contrast focused on a 2×2 subtable of the overall design. In practice, researchers rarely are content to make statements only at the omnibus level. Usually, more focused questions are pursued that turn to single degree-of-freedom interaction contrasts.

We can now make explicit why simple main effects do not elucidate the dynamics of statistical interaction. Interaction contrasts formally compare the effect of an independent variable on a dependent variable at one level of the moderator variable with that at another level of the moderator variable. By contrast, simple main effects make no such comparison. A simple main effect focuses on only one level of the moderator and asks if the independent variable has an effect at that particular level. For example, is there a gender difference for just Catholics? But the test of significance of this simple effect does not compare the effect with any other group (i.e., it does not compare the effect with that for Protestants or Jews). It simply does not address the issue of statistical interaction.

We can illustrate the point another way using a correlation example. Suppose that two variables, X and Y, are correlated .24 for males and .22 for females. Suppose that the correlation is statistically significant for males ($p < .05$) but not for females ($p > .05$). The significance tests within each group are analogous to simple effect tests. Can we conclude from these tests that the correlation between X and Y is stronger for males than it is for females? Of course not. Even though the correlation is statistically significant in one group but not in the other group, we can only say that the correlations differ if we directly test the difference between the two correlations. This test of the difference between the two correlations (which in this case is not statistically significant) is analogous to a test of interaction in the moderator framework.

In sum, our discussion of interaction analysis in later chapters will consider simple effects (which also are called conditioned effects), single degree-of-freedom interaction contrasts, and omnibus tests of interaction. All three types of tests potentially are of conceptual interest and all manifest themselves in interaction models using multiple regression. Although we have used qualitative variables and a simple factorial design to illustrate these concepts, we will see their counterparts in regression models with continuous predictors.

A Review of Multiple Regression

This section assumes that the reader is familiar with the basics of multiple regression. The intent is to introduce terminology and a frame of reference from which future discussions follow. For useful introductions to multiple regression, see Berry and Feldman (1985), Cohen and Cohen (1983), Lewis-Beck (1980), or Schroeder, Sjoquist, and Stephan (1986). As is traditional, we will use Greek notation and letters to refer to population coefficients and population data and Arabic letters to refer to sample estimates of the coefficients and sample data.

The Linear Model

Consider the case of three continuous variables, where the investigator is interested in the effects of two independent variables, X_1 and X_2, on a dependent variable, Y. The analysis in a set of sample data typically takes the form of a least squares regression equation such that

$$Y = a + b_1X_1 + b_2X_2 + e$$

where a is the least squares estimate of the population intercept, b_1 and b_2 are the least squares estimates of the population regression coefficients for X_1 and X_2, respectively, and e is a residual term. In this approach, several assumptions about the structure of the population data are required in order to apply ordinary least squares (OLS) analysis and its inferential tests in a strict sense: (1) The linear model being tested is the true model for all members of the population, (2) the residual terms in the population are independently distributed with a mean of zero, (3) the predictor variables are fixed in nature and have positive variance, (4) the rank of the sample data matrix equals the number of columns and is less than the number of observations (i.e., there is not complete multicollinearity), and (5) the residuals at a given fixed set of values of the X are normally distributed and have a variance that is equal to the residual variance at any other fixed set of values of X. When these assumptions are satisfied, an OLS estimator is said to be the best linear unbiased estimator (BLUE) in that it is linear, unbiased, and has minimum variance in the class of all linear unbiased estimators. Inferential tests are straightforward when these assumptions are met. Some of these assumptions can be relaxed with minor consequences to inferential tests or parameter estimation, whereas other assumption violations are problematic. For example, although many research applications rely on cases where the predictors are stochastic rather than fixed, OLS remains a viable approach if one assumes stochastic predictors conditional on the actual sample of observed Xs. We discuss issues of assumption violation in greater depth in Chapter 4.

The sample multiple correlation coefficient, R, is an index of overall model fit in the sample, and the regression coefficients often are interpreted as the "effect" of an X variable on Y partialing out all other X variables in the equation. Specifically, a given b represents the number of units that Y is predicted to change given a one-unit increase in X while "holding all other Xs constant." If all variables are standardized, then the intercept always equals zero and the b represent *standardized* regression coefficients. The interpretation of these coefficients is the same except that all units are expressed in terms of standard scores. Thus, a standardized b value of 1.5 for a predictor variable X implies that for every one standard score that X increases, Y is predicted to change 1.5 standard scores. Some social scientists prefer the use of standardized scores to unstandardized scores in multiple regression analysis because all of the variables are then thought to have a common metric and it supposedly is easier to make substantive comparisons of the magnitude of the coefficients for different independent variables. We discuss this issue in more detail in Chapter 4.

Consider the following example: A sociologist is interested in the extent to which overall satisfaction with one's marital relationship can be predicted from satisfaction or dissatisfaction with six distinct components of the relationship. Three hundred thirty-nine individuals rate how satisfied or dissatisfied they are with their overall marital relationship, using an 11-point, -5 to $+5$, rating scale ranging from very dissatisfied to very satisfied (with higher numbers indicating greater satisfaction). In addition, the individuals rate (on an identical scale) how satisfied they are with the following six aspects of their relationships: the amount of communication, the way affection is expressed, the amount of emotional support, the level of shared interests, the amount of time spent together, and the way conflict is resolved. The investigator performs a multiple regression analysis in which overall marital satisfaction is regressed onto the six components. Table 1.2 presents the results of the analysis using abbreviated versions of a computer printout from SPSS.

The squared multiple correlation was .663, which indicates the proportion of variance in the ratings of overall marital satisfaction that could be accounted for by the linear combination of the six component parts in the sample data. The null hypothesis that the population multiple correlation equals zero is tested by means of an F test, reported underneath the first table. The F is statistically significant [$F(6, 332) = 108.70$], leading us to reject the null hypothesis. The standardized and unstandardized regression coefficients are presented in the lower part of Table 1.2. As noted above, an unstandardized coefficient reflects the number of units that overall satisfaction is predicted to change given a one-unit increase in the X variable in question, holding all other X variables constant. For example, for every one rating scale unit that satisfaction with the amount of emotional support increases, the overall satisfaction with the marital relationship is predicted to change 0.307 rating scale units, holding all other satisfaction variables constant. By contrast, a one-unit increase in satisfaction with the amount of time spent together is associated with only a 0.005-unit predicted increase in overall satisfaction, holding all other components constant. The standardized coefficients are subject to the same form of interpretation but in terms of standard scores rather than raw scores.

TABLE 1.2
Abbreviated SPSS Ouput for Multiple Regression Example

Model	*R*	R^2	*Adjusted* R^2 *Square*	*SE of the Estimate*
1	.814	.663	.457	.74094

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	357.897	6	59.650	108.698	.000
Residual	182.191	332	.549		
Total	540.088	338			

	Unstandardized Coefficients		*Standardized Coefficients*			*95% Confidence Interval for B*		*Correlations*	
Model	*B*	*SE*	β	*t*	*Sig.*	*Lower Bound*	*Upper Bound*	*Zero-Order*	*Part*
(Constant)	1.010	.127		7.980	.000	.761	1.259		
Conflict	.121	.025	.191	4.880	.000	.072	.170	.578	.155
Time Together	.005	.023	.008	.220	.823	–.044	.054	.353	.007
Affection	.151	.039	.173	3.910	.000	.075	.227	.629	.210
Share Interest	.153	.034	.183	4.460	.001	.086	.220	.550	.142
Communication	.125	.037	.147	3.330	.000	.058	.192	.625	.106
Emotional Support	.307	.041	.350	7.560	.000	.227	.387	.724	.241

Each unstandardized regression coefficient has associated with it an estimated standard error (see the column labeled "Std. Error"). These statistics represent estimates of how much sampling error is operating when estimating the regression coefficients in the population. More specifically, the estimated standard error indicates (roughly speaking) the average deviation of sample estimates from the true value of the population parameter across all possible random samples of size *N*. The larger the standard error, the greater the amount of sampling error that is operating, everything else being equal, and the less confidence we have in the accuracy of the sample estimate.

The column labeled "*t*" indicates the *t* test for the null hypothesis that a given regression coefficient equals zero. The value of *t* equals the value of the regression coefficient divided by its estimated standard error. The column labeled "Sig." presents the *p* value for the *t* statistic. In this example, all the regression coefficients are statistically significant ($p < .05$) except the one associated with satisfaction about the time spent together.

Finally, additional insights into the relationship of each predictor to the criterion can be gained by examination of the zero-order correlations between each predictor and the criterion (see column labeled "Zero-Order") and the semipartial correlations between a given predictor and the criterion with all remaining predictors partialed out of the predictor (see the column labeled "Part"). The former statistic, when squared, reflects the proportion of explained variance in overall marital satisfaction that is accounted for by a given predictor when all other components of satisfaction are free to vary. The latter statistic, when squared, indicates the proportion of variance in overall marital satisfaction that is *uniquely* associated with a given component of satisfaction holding constant all other components. For example, satisfaction with emotional support accounts for $(100)(0.72)(0.72) = 51.5\%$ of the variance in overall marital satisfaction. This predictor uniquely accounts for $(100)(0.24)(0.24) = 5.8\%$ of the variance in marital satisfaction over and above that which is accounted for by the other five predictors.[1]

Hierarchical Regression

Often, researchers perform hierarchical multiple regression. In these cases, the investigator is interested in whether adding one or more predictor variables to an existing regression equation will significantly increase the predictability of the criterion. The amount of incremental explained variance is typically evaluated by subtracting the squared multiple correlation in the original equation from the squared multiple correlation in the expanded equation. The difference in the squared multiple correlations is the amount of incremental explained variance due to the additional predictors. For example, if the difference equals .10, then an additional 10% of explained variance in the criterion has resulted by the inclusion of the additional predictors. A test of the null hypothesis that the increment in the squared multiple correlation is zero in the population is yielded by the following equation:

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)} \qquad [1.1]$$

where R_2 is the multiple R for the expanded equation, R_1 is the multiple R for the original equation, k_2 is the number of predictors in the expanded equation, k_1 is the number of predictors in the original equation, and N is the total sample size. The resulting F is distributed with $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom.

The hierarchical test often is framed in terms of the effects of adding predictors to a base equation, as above. However, sometimes it is framed in terms of deleting predictors from an equation. Suppose we have a regression equation with six predictor variables and we want to determine the effect on the squared multiple correlation if we drop two of them from the model. The change in the squared Rs between the two equations reflects the proportion of explained variation that is lost by dropping the two predictors. A test of the statistical significance of the drop in explained variance uses Equation 1.1, where R_2 is the multiple correlation for the model with the greater number of predictors and R_1 is the multiple correlation for the model with the fewer number of predictors.

Categorical Predictors and Dummy Variables

Regression analysis often includes categorical variables as predictors, such as gender, ethnicity, and religious affiliation. Such variables are represented in the equation using dummy variables. A dummy variable is a variable that is created by the analyst to represent group membership. For example, in the case of gender, we can create a dummy variable and assign a 1 to all males and a 0 to all females. This method of scoring is called "dummy coding" and involves assigning a 1 to all members of one group and a 0 to everyone else. When a qualitative variable has more than two levels, it is necessary to specify more than one dummy variable to capture membership in the different groups. In general, one needs $m - 1$ dummy variables, where m is the number of levels of the variable. Suppose we had as a predictor variable a person's party affiliation that could take on three values, Democrat, Republican, or Independent. In this case, we need $3 - 1 = 2$ dummy variables to represent party affiliation. For the first dummy variable, D_D, we assign all Democrats a 1 and everyone else a 0. For the second dummy variable, D_R, we assign all Republicans a 1 and everyone else a 0. Although we could create a third dummy variable for Independents and assign them a 1 and everyone else a 0, such a variable is completely redundant with the other two dummy variables. Once we know whether someone is a Democrat and whether someone is a Republican (by means of the first two dummy variables), we know by definition whether he or she is an Independent. The reasoning behind this is more evident if one considers a dummy variable for gender. We create a single dummy variable to discriminate the two groups, whereby males are assigned a score of 1 and females a score of 0. If we create a second dummy variable that assigns a score of 1 to females and a score of 0 to males, it is perfectly negatively correlated with the first dummy variable and hence redundant. With dummy coding, the group that does not receive a 1 on any of the dummy variables is called the *reference group* for that variable. In the examples above, the reference group for gender is females and for party affiliation the reference group is Independents. The choice of which group is the reference group is arbitrary from a statistical point of view.

Suppose that we regress a measure of attitudes toward abortion onto a categorical variable of religion. The attitude variable ranges from 0 to 100, with higher scores indicating more positive attitudes. The religion variable has three groups, Catholic, Protestant, and Jewish. It is represented by two dummy variables, one in which all Catholics receive a 1 and everyone else receives a 0 (D_C) and the other where all Protestants receive a 1 and everyone else receives a 0 (D_P). The Jewish respondents are the reference group. Suppose that the regression analysis yielded a squared multiple correlation of .30. This indicates that religion accounts for 30% of the variance in the attitudes toward abortion. The F test of significance for the multiple correlation tests the null hypothesis that religion has no effect on attitudes toward abortion in terms of mean differences between the three groups. The unstandardized regression coefficients for the two dummy variables are meaningfully interpreted. Each coefficient reflects a mean difference. Specifically, the coefficient is the mean difference between the group scored 1 on the dummy variable minus the reference group. Suppose that the regression coefficient for D_C was -1.0. This indicates that the mean for Catholics minus the mean for Jews is -1.0, or that Catholics, on average, have an attitude toward abortion that is 1.0 unit lower than that of Jews. The regression coefficient for D_P was $-.5$, indicating that Protestants, on average, have an attitude toward abortion that is one half a unit lower than that of Jews. The significance tests for the regression coefficients are tests of significance of the mean difference in question.

If the above analysis was repeated but a continuous predictor was added to the regression equation, then the interpretation of the regression coefficients associated with the dummy variables still focuses on mean differences, but they are mean differences holding constant the other variables (i.e., the continuous variable) in the model. For example, if a measure of social class was added to the equation and the regression coefficient for D_C became $-.8$, then this is the predicted mean difference between Catholics and Jews holding constant or covarying out social class.

There are many ways in which scores can be assigned to dummy variables. As noted, we used a method called "dummy scoring" that relies on ones and zeros. Hardy (1993) discusses the logic of different coding schemes. Dummy coding is useful because it maps well onto the moderator conception of interaction effects, as will be shown in later chapters.

Predicted Values in Multiple Regression

Suppose a set of data is analyzed in which dollar contributions to an environmental group promoting the cleanup of a river are predicted from gender and a measure of ideology that reflects conservatism-liberalism. The ideology measure ranges from -3 to $+3$, with zero representing a neutral point and increasingly negative scores representing greater levels of conservatism and increasingly positive scores representing greater levels of liberalism. Gender is represented by a dummy variable, with males scored 1 and females scored 0. The outcome measure is in units of dollars. Suppose that the analysis yielded the following regression equation:

$$Y' = 10.00 + -5.50Gender + 1.00Ideology \tag{1.2}$$

where Y' is the predicted amount of money that the individual donates. We can calculate a predicted value of Y for any given profile of predictor variables by substituting the values for the predictors into the equation. For example, the predicted donation for males who have ideology scores of $+2$ is

$$Y' = 10.00 + -5.50(1) + 1.00(2) = 6.50$$

What is the predicted donation for males who have an ideology score of −2? By substitution, we obtain

$$Y' = 10.00 + -5.50(1) + 1.00(-2) = 2.50$$

We will make use of such predicted values in later expositions.

Transformations of the Predictor Variables

It is possible to perform algebraic manipulations on the predictor variables prior to performing a regression analysis to force the coefficients to reflect parameters that are of theoretical interest. The utility of doing so will be illustrated in later chapters, but we establish the basic logic here. Suppose that, prior to conducting the regression analysis in the previous example, we subtract a constant of 1 from the ideology scale. Whereas the original scale ranged from −3 to +3, the new scale ranges from −4 to +2, as each score is shifted down one unit. The results for the regression analysis using this transformed score are as follows:

$$Y' = 11.00 + -5.50Gender + 1.00Ideology_t$$

Note that the only parameter affected by this transformation is the intercept, with the other coefficients being identical to the original analysis. The intercept is the predicted value of Y when gender is 0 and when the transformed ideology value is 0. But a 0 on the transformed ideology variable represents a +1 on the original ideology variable. The intercept in this second analysis should equal the predicted value for females who have an ideology score of +1 in the original analysis. This is indeed the case. In the original equation,

$$Y' = 10.00 + -5.50(0) + 1.00(1) = 11.00$$

which is the same as the intercept in the second analysis. Why would one want to perform such transformations? Almost all computer packages report not only the parameter estimates for a regression equation but also the estimated standard errors and confidence intervals for a given estimate. Using transformations such as the above represents a simple (but cumbersome) way for calculating the confidence interval for the predicted mean Y value of any predictor profile. Simply transform each predictor by adding or subtracting a constant so that a score of zero on the transformed variable represents the predictor value on the original scale that you are interested in. The intercept term from the equation using the transformed predictors will then provide the predicted mean Y value for that particular profile, and the confidence interval for the predicted mean will be those associated with the intercept term. In the absence of transformations of this nature, the intercept term sometimes has limited interpretational value because it may reflect the predicted mean Y for the case where values of zero on the predictors are nonexistent or outside of the range of the values being studied. We make use of the transformation strategy and variants of it in later chapters, and it will be a key strategy used to isolate a variety of significance tests and confidence intervals.

Overview of the Book

In the remainder of this book, we discuss a range of issues pertinent to interaction analysis. Our focus is primarily on the analysis of continuous predictor variables and mixtures of continuous and qualitative predictor variables because excellent treatments already exist for the case of all qualitative predictor variables (e.g., Cohen & Cohen, 1983). In Chapter 2, we focus on two-way interactions and introduce traditional product term analysis as a means of analyzing bilinear interaction effects. In Chapter 3, we consider three-way interactions. Chapter 4 focuses on assorted topics that have emerged in the formal literature on interaction analysis and that will assist the applied researcher in thinking about fundamental issues when using interaction models.

2. TWO-WAY INTERACTIONS

In this chapter, we first consider general issues with the specification of interaction models in multiple regression. We then consider the analysis of two-way interactions where both of the predictor variables are continuous in form. Finally, we discuss the case where one of the predictors is qualitative and one is continuous.

Regression Models With Product Terms

Consider the case of three continuous variables, where the investigator is interested in the effects of two independent variables (X and Z) on a dependent variable (Y). To use a concrete example, suppose that an investigator is interested in understanding why some teenagers engage in sex without using birth control although other teenagers tend to use birth control. A sample of 125 sexually active female teenagers is studied, and for each teen, a measure of their intention to use birth control is obtained. The measure consists of a rating scale with endpoints "definitely do not intend to use birth control" to "definitely intend to use birth control." Scores can range from 0 to 30, with higher scores indicating a stronger intention to use birth control. The researcher hypothesizes two classes of factors that influence this intention. The first is the individual's personal feelings or attitude toward using birth control, that is, how favorably or unfavorably the individual personally feels about birth control. The second factor is the perceived peer pressure to use birth control. Each of these factors, attitudes and perceived peer pressure, is measured on a 5-point scale ranging from 1 to 5. For the attitude measure, the higher the score, the more favorably the individual personally feels about using birth control. For the peer-pressure measure, low scores imply relatively little peer pressure to use birth control and high scores imply increasingly higher levels of pressure.

As noted in Chapter 1, the test of an additive (or "main-effects") model for predicting Y from X and Z typically takes the form of a least squares regression equation, where the population model is

$$Y = \alpha + \beta_1 X + \beta_2 Z + \varepsilon \qquad [2.1]$$

Application of Equation 2.1 involves regressing the measure of intention onto the measures of attitudes and peer pressure, respectively. Suppose that the investigator was interested in exploring the presence of an interaction effect. Specifically, the researcher hypothesized that the relationship between attitudes and intention is moderated by the amount of peer pressure that is present: When peer pressure is minimal, personal attitudes will exert a strong effect on intentions. However, when peer pressure is strong, the influence of attitudes will be less. In this case, the focal independent variable is the attitude toward using birth control (X) and the moderator variable is peer pressure (Z).

The most common approach to modeling interactions in regression analysis is to use product terms. We can illustrate the basic logic of product terms in interaction models by simple algebra. If the effect of attitudes on intentions is reflected by β_1 in Equation 2.1 and if Z is thought to moderate the effect of attitudes on intentions, then it follows that as Z varies, the value of β_1 also should vary. One way of expressing how β_1 might vary as a function of Z is in terms of a linear function:

$$\beta_1 = \alpha' + \beta_3 Z \qquad [2.2]$$

According to this formulation, for every one unit that Z changes, the value of β_1 is predicted to change by β_3 units. We now substitute the right-hand side of the above expression for β_1 in Equation 2.1, yielding

$$Y = \alpha + (\alpha' + \beta_3 Z)X + \beta_2 Z + \varepsilon$$

Multiplying this out yields

$$Y = \alpha + \alpha' X + \beta_3 XZ + \beta_2 Z + \varepsilon$$

and after assigning new labels to the coefficients and rearranging terms, we obtain an interaction model with a product term:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon \qquad [2.3]$$

This exposition is an oversimplification because other types of interaction models lead to the same equation and other equation forms evolve from different interaction models. Our goal here is merely to provide the reader with a sense of the rationale behind the use of a product term to reflect an interaction where the effect of the focal independent variable on the outcome variable is said to be a linear function of the moderator variable.

Two Continuous Predictors

The Traditional Regression Strategy

We illustrate crucial points in the analysis of interaction effects using the prior example predicting intentions from attitudes and peer pressure. The data for this example are in Table 2.1. They are hypothetical and have been constructed to have unrealistically systematic properties in order to illustrate later points in our discussion. For pedagogical purposes, the data can be arranged into a factorial table, with mean Y scores represented at each combination of X and Z. This has been done in Table 2.2. This table is essentially a 5×5 factorial design.

The most popular approach to analyzing the interaction effect in multiple regression uses a strategy recommended by Cohen and Cohen (1983). It involves forming the product term, XZ, which is said to encompass the interaction effect, and then to calculate two R^2 values, one for the "main-effect-only" model (Equation 2.1) and another that adds the product term to the main-effect model, yielding the three-predictor equation for a set of sample data:

$$Y = a + b_1 X + b_2 Z + b_3 XZ + e \qquad [2.4]$$

If an interaction effect is present, then the difference between the two R^2 values should be statistically significant (barring a Type I error). The formal significance test of this difference uses the hierarchical F test of Equation 1.1.

TABLE 2.1
Hypothetical Data for Interaction Example

ID	*Y*	*X*	*Z*	*ID*	*Y*	*X*	*Z*	*ID*	*Y*	*X*	*Z*
1	3	1	1	43	7	2	4	85	19	4	2
2	4	1	1	44	8	2	4	86	12	4	3
3	5	1	1	45	9	2	4	87	13	4	3
4	6	1	1	46	4	2	5	88	14	4	3
5	7	1	1	47	5	2	5	89	15	4	3
6	3	1	2	48	6	2	5	90	16	4	3
7	4	1	2	49	7	2	5	91	9	4	4
8	5	1	2	50	8	2	5	92	10	4	4
9	6	1	2	51	13	3	1	93	11	4	4
10	7	1	2	52	14	3	1	94	12	4	4
11	3	1	3	53	15	3	1	95	13	4	4
12	4	1	3	54	16	3	1	96	6	4	5
13	5	1	3	55	17	3	1	97	7	4	5
14	6	1	3	56	11	3	2	98	8	4	5
15	7	1	3	57	12	3	2	99	9	4	5
16	3	1	4	58	13	3	2	100	10	4	5
17	4	1	4	59	14	3	2	101	23	5	1
18	5	1	4	60	15	3	2	102	24	5	1
19	6	1	4	61	9	3	3	103	25	5	1
20	7	1	4	62	10	3	3	104	26	5	1
21	3	1	5	63	11	3	3	105	27	5	1
22	4	1	5	64	12	3	3	106	19	5	2
23	5	1	5	65	13	3	3	107	20	5	2
24	6	1	5	66	7	3	4	108	21	5	2
25	7	1	5	67	8	3	4	109	22	5	2
26	8	2	1	68	9	3	4	110	23	5	2
27	9	2	1	69	10	3	4	111	15	5	3
28	10	2	1	70	11	3	4	112	16	5	3
29	11	2	1	71	5	3	5	113	17	5	3
30	12	2	1	72	6	3	5	114	18	5	3
31	7	2	2	73	7	3	5	115	19	5	3
32	8	2	2	74	8	3	5	116	11	5	4
33	9	2	2	75	9	3	5	117	12	5	4
34	10	2	2	76	18	4	1	118	13	5	4
35	11	2	2	77	19	4	1	119	14	5	4
36	6	2	3	78	20	4	1	120	15	5	4
37	7	2	3	79	21	4	1	121	7	5	5
38	8	2	3	80	22	4	1	122	8	5	5
39	9	2	3	81	15	4	2	123	9	5	5
40	10	2	3	82	16	4	2	124	10	5	5
41	5	2	4	83	17	4	2	125	11	5	5
42	6	2	4	84	18	4	2				

TABLE 2.2
Cell Means as a Function of X and Z

	Peer Pressure (Z)				
Attitude (X)	*1*	2	*3*	*4*	*5*
5	25	21	17	13	9
4	20	17	14	11	8
3	15	13	11	9	7
2	10	9	8	7	6
1	5	5	5	5	5

Analyzing the data in Table 2.1 via standard multiple regression procedures without the product term yields a multiple R for the two-term additive model of .90139. The regression equation is

$$Y = 8.0 + 3.0X + -2.0Z + e \tag{2.5}$$

The multiple R for the three-term interaction model is .96825, and the regression equation is

$$Y = -1.0 + 6.0X + 1.0Z + -1.0XZ + e \tag{2.6}$$

Applying Equation 1.1 yields the following:

$$F = \frac{(0.96825^2 - 0.90139^2)/(3 - 2)}{(1 - 0.96825^2)/(125 - 3 - 1)} = 242.26$$

For 1 and 121 *df*, the F is statistically significant, implying the presence of statistical interaction. The hierarchical F test yields the same p value as that of the t test for the statistical significance of b_3 in Equation 2.4. For our data, the t for b_3 was 15.56. The square of this is the same as the observed F, namely (15.56)(15.56) = 242.26. Thus, instead of conducting the hierarchical test to determine statistical significance of the interaction effect, one simply can examine the significance test for b_3.

We now consider selected issues central to this analysis. We use the birth-control example to discuss these issues. Afterward, we apply the concepts to a new example in order to illustrate typical analytic strategies in practice.

The Form of the Interaction

As noted, simple product terms permit the investigator to test for the presence of a moderated relationship. In principle, there is a wide variety of moderated relationships that can characterize an interaction effect between continuous variables. For example, one functional form is when the slope between Y and X changes as a linear function of scores on Z. This is called a *bilinear interaction*, and the data in Table 2.2 have this form. When peer pressure is low, equaling a score of 1, the slope of intentions on attitudes is high, namely 5.0: For every one unit that attitudes change, intentions are predicted to change five units. As peer pressure increases, the slope decreases. For example, when peer pressure equals 3, the slope of intentions on attitudes equals 3.0: For every one unit that attitudes change, intentions are predicted to change three units. Note that there is a simple linear relationship between changes in peer pressure and changes in the slope of intentions on attitudes. Every time peer pressure increases by one unit, the slope decreases by one unit. This orderly, monotonic, linear relationship between changes in the slope and changes in the moderator variable is the essence of statistical interaction as measured by traditional product terms.

Other types of functional forms are possible. For example, it might be the case that changes in the intention-attitude slope are relatively large as one moves from low peer pressure to moderate amounts of peer pressure. However, as one progresses from moderate to high peer pressure, the changes in slopes may become less dramatic until they reach a point of minimal change. Alternatively, the relationship between intentions and attitudes may be nonlinear in form. The *shape* (rather than the slope) of the curve between the independent variable and dependent variable may change as a function of peer pressure. The number of possible functional forms of moderated relationships involving continuous predictor variables is infinite. An important point to remember is that simple product terms as used in Equation 2.4 test for only one functional form, namely, a bilinear interaction. Failure to obtain a statistically significant interaction using traditional product terms may reflect the presence of an alternative functional form rather than the absence of a moderated relationship. In Chapter 4, we discuss strategies for identifying alternative forms of interactions and strategies for testing them. In the ideal situation, a theory will predict the presence of a specific functional form and then the data analyst will construct the appropriate model to test for that functional form. In the remainder of this section, we restrict our attention to the analysis of bilinear interactions.

Interpreting the Regression Coefficient for the Product Term

All of the regression coefficients yielded by Equation 2.4 are subject to meaningful interpretation. We first consider b_3, the regression coefficient associated with the product term. b_3 is a single-degree-of-freedom interaction contrast that conveys information about the nature of the interaction. It indicates the number of units that the slope of Y on X changes given a one-unit increase in Z. In our example, the value of b_3 is -1.0. This means that for every one unit that peer pressure increases, the slope of intentions on attitudes is predicted to decrease by one unit. That this is indeed the case is apparent from inspection of Table 2.2.

The data in Table 2.2 are hypothetical, and the slope of Y on X at a given value of Z is evident from visual inspection of the data. However, rarely will data be so orderly in practice. As it turns out, it is possible to calculate the estimated effect of X on Y for any given value of Z using the regression coefficients from the three-term regression equation. All that is required is some algebraic manipulation.

We begin by specifying a value of Z (peer pressure) where we want to analyze the relationship between Y and X. Let us begin with the lowest score possible on Z, namely, a value of 1. Using the three-term equation, we substitute the score of 1 wherever Z occurs. This yields

$$Y = -1.0 + 6.0X + (1.0)(1) + (-1.0)(X)(1) + e$$

We can rearrange the right side of the equation to group all of the terms with an X to the right:

$$Y = -1.0 + (1.0)(1) + 6.0X + (-1.0)(X)(1) + e$$

Then we can factor X out of the relevant terms:

$$X = -1.0 + (1.0)(1) + [6.0 + (-1.0)(1)]X + e$$

which yields

$$Y = 0.0 + 5.0X + e$$

The result is the linear equation describing the relationship between Y and X when Z equals 1. Let us now perform the same calculations using the highest Z score of 5. Substituting a score of 5 for Z yields

$$Y = -1.0 + 6.0X + (1.0)(5) + (-1.0)(X)(5) + e$$

Performing the same calculations produces the following result:

$$Y = 4.0 + 1.0X + e$$

These calculations reveal how the relationship between Y and X varies across the range of Z: At low values of Z (i.e., 1), a one-unit increase in X is associated with a five-unit predicted increase in Y. At high values of Z (i.e., 5), a one-unit increase in X is associated with a one-unit predicted increase in Y. We can formally state an equation for calculating the slope of the predicted effects of X on Y at any particular value of Z. It is

$$b_1 \text{ at } Z = b_1 + b_3 Z \qquad [2.7]$$

To illustrate from our example, the slope of Y on X when Z equals 4 is

$$b_1 \text{ at } 4 = 6.0 + (-1.0)(4) = 2.0$$

If we were to calculate the values of the slope of Y on X at each of the five values of Z, we would find the following:

b_1 at a score of 1 on $Z = 5.0$
b_1 at a score of 2 on $Z = 4.0$
b_1 at a score of 3 on $Z = 3.0$
b_1 at a score of 4 on $Z = 2.0$
b_1 at a score of 5 on $Z = 1.0$

Notice that for every one unit that Z increases, the value of the slope decreases by -1.0, which is the value of b_3. Again, the value of the product-term coefficient reveals how the slope is predicted to change given a one-unit change in the moderator variable.

Interpreting the Regression Coefficients for the Component Terms

Some social scientists have argued that including a product term in an equation yields regression coefficients for the component parts (X and Z) that are difficult to interpret. Researchers have noted that these coefficients often change dramatically when compared with the corresponding coefficients from a "main-effects-only" model. Indeed, the coefficients may even reverse themselves in sign. Such "contradictions" supposedly make interpretation of the regression coefficients difficult in a model with product terms.

In actuality, the regression coefficients yielded in the three-term equation are subject to meaningful interpretation. Differences such as those noted above occur because the coefficients in the two equations estimate different concepts. In the two-term "main-effects-only" model, a regression coefficient estimates the effect of a predictor variable on the dependent variable taking into account each level of the other predictor variables. Speaking somewhat informally, in the two-term model, b_1 reflects the trends of change in Y given a unit change in X at each level of Z; b_2 reflects the trends of change in Y given a unit change in Z at each level of X. In contrast, for the model with a product term, the regression coefficients for X and Z reflect *conditional* relationships for a specific value of Z or X: b_1 reflects the influence of X on Y when Z equals 0, and b_2 reflects the influence of Z on Y when X equals 0. Differences in the values of the b_1 and b_2 coefficients in the two models result from the fact that in the "main-effects" model, the coefficients estimate "general" relationships averaging across the levels of the other predictor variable, whereas in the product-term model, they estimate conditional relationships focused on a specific value of the other predictor involved in the product term.

Stated another way, in the product-term model, the regression coefficients for X and Z do not represent main effects, as they are traditionally thought of. Rather, the coefficients represent simple effects. The coefficient for X estimates the effect of X on Y when Z is at a specific value, namely, when $Z = 0$. The coefficient for Z estimates the effect of Z on Y when X is at a specific value, namely, when $X = 0$. Some researchers mistakenly interpret these coefficients as if they were main effects. They are not. It is important to keep in mind the nature of these coefficients.

For the example on birth control, the coefficient for X represents the estimated effect of attitudes on intentions when peer pressure corresponds to a Z value of 0. The value of the coefficient is 6.0, so that when peer pressure corresponds to a score of 0 on Z, a one-unit change in attitude is predicted to yield a six-unit change in intention.

However, there is a problem with this interpretation. The coefficient reveals the estimated effect of X on Y when $Z = 0$, but a score of 0 on the measure of peer pressure does not exist! The values of Z range from 1 to 5, so characterizing the effect of attitudes on intentions when $Z = 0$ does not make substantive sense. One solution to this problem is to transform Z so that a 0 value is meaningful. For example, the mean score on Z is 3.0. Suppose we subtract a value of 3.0 from the Z score of each individual. Instead of ranging from 1 to 5, the transformed Z variable, Z_t, will range from −2 to +2. Instead of the mean value of 3.0 on Z, the mean value for Z_t will be 0. The act of subtracting the mean from a variable to form a new scale for that variable is called *mean centering*. Suppose we mean center Z, recalculate the product term by forming XZ_t, and then estimate the regression equation:

$$Y = a + b_1X + b_2Z_t + b_3XZ_t + e \qquad [2.8]$$

The original equation using Z as the predictor was

$$Y = 1.0 + 6.0X + 1.0Z + -1.0XZ + e$$

The regression equation when Z_t is used in place of Z is

$$Y = 2.0 + 3.0X + 1.0Z_t + -1.0XZ_t + e$$

There are several features of the two analyses worth noting. First, the squared multiple correlation predicting Y from X, Z, and XZ and its significance test is identical to that for predicting Y from X, Z_t, and XZ_t. The multiple correlation (and its significance test) is invariant to the transformation that we performed. Second, the value of b_3 and its significance test also is invariant to the transformation. Exactly the same characterization of the interaction effect will result in both analyses. Note, however, that the coefficient for X has changed from 6.0 in the original analysis to 3.0. In the new analysis, the coefficient for X is the predicted effect of X on Y when $Z_t = 0$. In this analysis, a score of 0 on Z_t corresponds to a score of 3 on the untransformed Z. So the coefficient actually reflects the effect of X on Y when $Z = 3$. In other words, when peer pressure is moderate, as reflected by a score of 3 on Z, the effect of X on Y as reflected by the slope of Y on X is 3.0. By transforming the value of Z so that the transformed variable score of 0 is meaningful, the coefficient associated with X becomes meaningful. In the new analysis, b_1 is the predicted effect of X on Y when Z equals its sample mean.

We now have two ways of calculating the effect of X on Y at a given value of Z for an interaction model. One strategy is to use Equation 2.7, which we repeat here:

$$b_1 \text{ at } Z = b_1 + b_3Z$$

For example, the value of b_1 when $Z = 2$ is $6.0 + 1.0(2) = 4.0$. The second strategy is to transform Z so that the score of 0 on Z_t corresponds to the value of interest on the original metric, form the product term using Z_t, and then calculate the regression equation accordingly. For example, if we subtract 2 from each value of Z, then Z_t ranges from -1 to 3 and a score of

0 corresponds to a score of 2 on the original metric. The value of b_1 in the regression of Y onto X, Z_t, and XZ_t will yield a b_1 value of 4.0. Although this latter strategy seems cumbersome as a way of estimating the effect of X on Y at a specific value of the moderator variable, it has advantages that manifest themselves in later sections.

Significance Tests and Confidence Intervals

Multiple regression programs in all of the major statistical computer packages routinely provide estimated standard errors, significance tests, and confidence intervals for the regression coefficients in the model. The significance test for b_3 is a test of the null hypothesis of no bilinear interaction. As noted earlier, this t test yields the same p value as that of the more traditional hierarchical F test for adding a product term to a main-effect model, so it is not necessary to conduct the hierarchical analysis for this purpose.

For the "main-effect" terms of the interaction model, the conditional nature of the regression coefficients also applies to the estimated standard errors of the coefficients. The estimated standard errors for the regression coefficients associated with X and Z in the interaction model are conditional and reflect sampling error when the other predictor in the product term equals 0. Thus, the standard error for b_1 in Equation 2.4 estimates sampling error for the regression coefficient of Y on X when Z equals 0. Similarly, the standard error for b_2 in Equation 2.4 estimates sampling error for the regression coefficient of Y on Z when X equals 0.

In the previous section, we showed how one could calculate the slope of Y on X for any given value of Z by using Equation 2.7. It also is possible to calculate an estimated standard error for this coefficient. From the equation that includes the product term, the estimated standard error is

$$SE(b_1 \text{ at } Z) = [(\text{var}(b_1) + Z^2 \text{var}(b_3) + 2Z \,\text{cov}(b_1, b_3)]^{1/2} \qquad [2.9]$$

where $\text{var}(b_1)$ is the variance of the b_1 regression coefficient, $\text{var}(b_3)$ is the variance of the b_3 regression coefficient, and $\text{cov}(b_1, b_3)$ is the covariance of the b_1, b_3 regression coefficients. The variance and covariance terms on the right-hand side of the equation are obtained from standard computer output, although these values are typically not reported as default options. The significance test of b_1 at a given value of Z takes the form of a t test that divides the coefficient by its estimated standard error, such that

$$t = (b_1 \text{ at } Z)/SE(b_1 \text{ at } Z) \qquad [2.10]$$

where Z is the value of Z at which the effect of X on Y is to be tested. The t value in Equation 2.10 is distributed as t with $N - k - 1$ degrees of freedom, where k is the number of predictor terms in the interactive model (in this case, $k = 3$).

A far easier strategy to calculating the relevant estimated standard error, confidence interval, and significance test for b_1 at any given value of Z is to use the transformation strategy discussed earlier. One performs a simple transformation of Z by subtracting a constant from it so that the score of 0 on Z_t corresponds to the original Z value of interest. Then use a computer program to regress Y onto X, Z_t, and XZ_t. The value of b_1 will be the coefficient for the slope of Y on X when $Z_t = 0$, and its estimated standard error, significance test, and confidence interval will all be the desired parameters for that conditional coefficient. We illustrate the application of these ideas in a later section.

To summarize thus far, the traditional interaction model involving two continuous predictors is tested in sample data using the equation

$$Y = a + b_1X + b_2Z + b_3XZ + e$$

The coefficient b_3 in this model is a single-degree-of-freedom interaction contrast and indicates by how many units the slope of Y on X is predicted to change given a one-unit change in the moderator variable, Z. The significance test of b_3 tests the null hypothesis of no bilinear interaction. The coefficient b_1 is a simple effect and reflects the effect of X on Y when $Z = 0$. A researcher may be interested in characterizing the effect that X has on Y at selected values of Z, and the transformation strategy can be used to do so. This strategy also yields relevant estimated standard errors, significance tests, and confidence intervals for these simple effects.

Multicollinearity

Some researchers are wary of interaction analysis with product terms because the product term often is highly correlated with the component parts used to define the product term. If XZ is highly correlated with either X or Z or both, the fear is that the evaluation of the interaction effect will be undermined due to problems of multicollinearity. This fear usually is misguided.

We noted earlier that the significance test of the interaction, that is, the significance test of b_3, is invariant to a simple transformation that subtracts a constant from Z, one that subtracts a constant from X, or one that subtracts constants from both X and Z. Although the value and t test of the interaction coefficient is unaffected by these transformations, the transformations do affect the correlations between XZ and X and between XZ and Z. Sometimes the transformations will increase these correlations, and sometimes they will decrease the correlations. If X and Z are normally distributed, then mean centering both X and Z prior to forming the product term will result in a product term that is uncorrelated with both X and Z (see Cronbach, 1987, for elaboration). Yet despite this, the result of the significance test of b_3 and the confidence intervals for b_3 are identical to the case where no transformation is made and the correlation of XZ with its component parts is substantial!

The statistical dynamics underlying this result need not concern us here, and interested readers are referred to Friedrich (1982) and Cronbach (1987). The major point is that high levels of collinearity between a product term and its component parts generally will not be problematic for interaction analysis unless the collinearity is so high that it disrupts the computer algorithm designed to isolate the relevant standard errors in a standard computer statistical package. If this turns out to be the case (as reflected by an error message in the computer output), one can simply mean center X and mean center Z, recalculate the product term using the mean-centered scores, and rerun the analysis. In most cases, this will reduce dramatically the collinearity and eliminate the computational problem.

Although high collinearity between XZ and X and between XZ and Z usually is not problematic, this is not true of collinearity between X and Z. High collinearity between X and Z can lead to serious complications.

Strength of the Interaction Effect

The strength of the interaction effect can be evaluated by numerous statistics, either in the form of unstandardized measures of effect size or standardized measures of effect size. The most popular unstandardized index is simply the value of b_3. As b_3 deviates from zero, the interaction effect is stronger, everything else being equal. The most popular standardized effect-size measure is the squared semipartial correlation for the product term holding constant its component parts. This value reflects the proportion of variance in the dependent variable that is accounted for uniquely by the interaction effect. It can be calculated by computing the difference in squared multiple correlations for the "main-effect-only" model as compared with the interaction model. In the birth-control example, the two-term additive model yielded a squared multiple correlation of .811, whereas the three-term interaction model yielded a squared multiple correlation of .937. The "strength" of the interaction effect was therefore $.937 - .811 = .126$. The interaction effect accounted for 12.6% of the variance in intentions to use birth control in the sample data. This index is positively biased, but the bias

TABLE 2.3
Means and Standard Deviations for Variables

Variable	*Mean*	*SD*
Desired family size	4.440	2.748
Family size in which raised	2.960	1.601
Income	34.933	14.220

tends to decrease with larger sample sizes and higher values of R^2. For potential pitfalls in the interpretation of standardized effect-size measures, see McClelland and Judd (1993) and Jaccard (1998).

A Numerical Example

At this point, a concrete example will help to summarize the major points of our discussion. In a sociological study, 100 religious Catholics from a Midwestern community were administered a scale measuring their intention to have a large family (Y). Values could range from 0 to 15, with higher numbers indicating stronger intent to have a large family. In addition, respondents were asked to indicate the number of children there were in the family in which they were raised as well as their current family income measured in units of \$1,000 (e.g., a score of 15 = \$15,000). The means and standard deviations for the three measures are in Table 2.3.

The researcher hypothesized that individuals who were raised in larger families would have more positive attitudes toward large families than individuals raised in smaller families and hence would have a stronger intent to have a large family. She also hypothesized that this effect of the family size in which the individual was raised on family-size intentions would be moderated by family income. Her logic was that when individuals are relatively poor, they will be less likely to translate their desires for a larger family into reality because of the costs of raising children. For wealthier families, such costs are not a constraint. Thus, the researcher expected to see a stronger relationship between the family size in which the individual was raised and the intent to have a large family when people were wealthy as opposed to when people were poor. In this study, the intention to have a large family is the outcome variable (Y), the size of the family in which one was raised is the focal independent variable (X), and income is the moderator variable (Z).

For the analysis, the X and Z scores were mean centered to avoid problems with multicollinearity and to make the b_1 and b_2 coefficients more interpretable. The product of the centered scores was computed for each respondent. A multiple regression analysis was then conducted using SPSS that regressed Y onto X, Z, and XZ (note: hereafter, X and Z refer to centered scores). The correlations between the product term and its centered component parts were trivial ($r = -.01$ and $.019$, respectively).

The multiple correlation for the interaction model was .725, and the regression equation was

$$Y = 4.4279 + 0.81324X + 0.0997Z + 0.0149XZ + e$$

The estimated standard errors for b_1, b_2, and b_3 were 0.098, 0.045, and 0.007.

The t test of the b_3 coefficient yielded a statistically significant result ($t = 2.187$, $p < .04$), suggesting the presence of an interaction effect. The 95% confidence interval for b_3 was 0.001 to 0.029. The strength of the interaction effect, as indexed by the squared semipartial correlation for the product term, was .017. This was calculated by taking the difference in squared multiple correlations for the "main-effects-only" model and the interactive model. For the former, the squared multiple correlation was .509, and for the latter, it was .526. This yields $.526 - .509 = .017$. The interaction effect accounts for 1.7% of the variance in desired family size.

The nature of the interaction effect is captured in the value of b_3. The value of b_3 indicates how the relationship between family size intentions (FSI) and family size in which one was raised (RAISE) varies across income. For every \$1,000 that income increases (which corresponds to "one unit" on Z), the slope of FSI on RAISE is predicted to increase 0.0149 units.

To provide the reader with an intuitive sense of how the slope of FSI on RAISE differs depending on the value of the moderator variable, we calculated a simple effect for predicting FSI from RAISE at three different values of income, a "low" income value, a "medium" income value, and a "high" income value. Let a "low" income be represented by an income that is 1 *SD* below the mean income, a "medium" income be represented by an income at the mean, and a "high" income be represented by an income that is 1 *SD* above the mean income. The standard deviation on income was 14.220. A "low" income corresponds to $34.933 - 14.220 = 20.713$, or \$20,713; a "medium" income corresponds to 34.933, or \$34,933; and a "high" income corresponds to $34.933 + 14.220 = 49.153$, or \$49,153. The slope of FSI on RAISE when income is "medium" is obtained directly from the output because the data for X and Z were mean centered. It corresponds to b_1 and equals 0.813. The estimated standard error for b_1 as reported in the output is 0.098 and the t value for the test of significance is 4.43 ($p < .05$). The 95% confidence interval for b_1 is 0.620 to 1.009. Although we could perform all the necessary calculations using the equations presented earlier to isolate the corresponding statistics for "low" and "high" income values, we used the transformation strategy. For the "low" income value, we subtracted 20.713 from the original income scores, multiplied this transformed score by the original X scores, and then regressed Y onto the relevant predictors (X, Z_t, and XZ_t). We then noted the values associated with b_1. We repeated this process using the "high" income score (34.933 + 14.220 = 49.153, or $49,153), in which the value 49.153 was subtracted from the original income scores. The results for b_1 for these analyses are summarized below:

Income Level	b_1	*SE*	*95% CI*	t	*p Value*
Low	0.602	0.136	0.333 to 0.873	4.43	<.001
Average	0.815	0.098	0.620 to 1.009	8.33	<.001
High	1.026	0.136	0.757 to 1.297	7.55	<.001

At "low" levels of income, each additional child in the family in which one was raised translates into an additional increase of 0.602 in one's intention to have a large family. At "medium" levels of income, each additional child in the family in which one was raised translates into an additional increase of 0.815 in one's intention to have a large family. At "high" levels of income, each additional child in the family in which one was raised translates into an additional increase of 1.026 in one's intention to have a large family. The effects of previous family size on the intention to have a large family were statistically significant at all three of these values. It can be seen from these analyses that the effects of RAISE get larger as income increases, which is in accord with the value of b_3 and the hypothesis of the investigator. These exemplar simple effects help provide readers with an intuitive sense of the interaction, but they do not represent a formal test of the interaction. The test of the interaction resides in the significance test for b_3.

Graphical Presentation

Some researchers like to convey interactions using graphs. One approach is to plot the three regression lines for the regression of Y on X at the "low," "medium," and "high" values of Z, as calculated above. The relevant slopes were 0.602, 0.815, and 1.027, respectively. To generate such plots, we also need the value of the intercept for each regression line. These can be calculated from the original regression equation using the following formula:

$$\text{intercept for } Y \text{ on } X \text{ at } Z = a + b_2 Z \qquad [2.11]$$

where a is the intercept in the regression equation for the interaction model and b_2 is the regression coefficient associated with the moderator variable

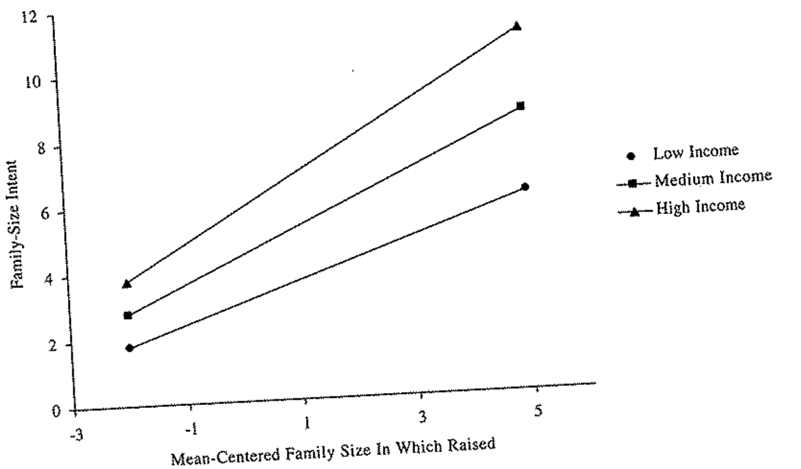

Figure 2.1 Regression Lines Predicting Family-Size Intent From Family Size in Which One Was Raised at Three Levels of Income

in that equation. Because we used a regression equation with mean-centered X and Z scores, the "low" score on Z is -14.22, the "medium" score is 0, and the "high" score is 14.22. Substituting the relevant values into Equation 2.11 yields

$$\text{intercept at } -14.22 = 4.4279 + 0.0997(-14.22) = 3.0102$$
$$\text{intercept at } 0 = 4.4279 + 0.0997(0) = 4.4279$$
$$\text{intercept at } 14.22 = 4.4279 + 0.0997(14.22) = 5.8456$$

Figure 2.1 presents three regression lines that conform to the three equations $Y = 3.0102 + 0.603X$, $Y = 4.4279 + 0.815X$, and $Y = 5.8456 + 1.026X$ for the "low," "medium," and "high" values of Z, respectively. Keep in mind that the X in these equations refers to the mean-centered X scores. If there were no interaction effect, the three lines would be parallel. It is evident that this is not the case.

A Qualitative Predictor and a Continuous Predictor

To illustrate the case of a qualitative and a continuous predictor, we use a hypothetical example where a researcher investigates the relationship between how satisfied adolescents are with their relationships with their mothers, the gender of the adolescent, and the amount of time that the mother and adolescent spend together. The outcome measure, adolescent satisfaction, was measured on a 21-point scale ranging from 0 to 20 and asked adolescents how satisfied they were with their relationship with their mother. Higher scores indicated higher levels of satisfaction. Gender of the adolescent was represented by a dummy variable scored as 1 = male and 0 = female. The amount of time the mother and child spent together was based on a self-report measure from the mother, who reported the number of hours the two spent together during a typical week. The mean number of hours spent together was 24.06, with a standard deviation of 4.62. The sample size was 100. We consider first the case where the number of hours spent together is conceptualized as the focal independent variable and gender is the moderator variable. Then we repeat the analysis reversing the role of the two predictor variables.

A Qualitative Moderator Variable

In this framework, we are interested in whether the effect of the amount of time the mother and adolescent spend together on adolescent relationship satisfaction is different for males and females. Stated more formally, we are interested in whether the regression coefficient when regressing adolescent satisfaction onto time spent together for males is different from the corresponding regression coefficient for females. If the two slopes are identical, then the effects of time spent together on adolescent satisfaction are the same for males and females and there is no interaction effect. However, if the slopes differ, then gender moderates the impact of time spent together on adolescent relationship satisfaction.

We first mean center the time predictor and then regress the satisfaction ratings (Y) onto the mean-centered self-reports of time spent together (X), the dummy variable for gender (Z), and the product term (XZ). Note that dummy variables are not centered. The squared multiple correlation was .784, which was statistically significant [$F(3, 97) = 116.07$, $p < .01$]. The regression equation was

$$Y = 7.645 + 0.247 Time_c + -5.270 Gender$$
$$+ -0.260(Time_c)(Gender) + e$$

The regression coefficient for the interaction term, –0.260, had an estimated standard error of 0.113 and a 95% confidence interval of –0.484 to –0.036. The t test of the coefficient was statistically significant ($t = 2.30$, $p < .05$), suggesting the presence of an interaction. The squared semipart correlation for the product term was .012, indicating that the interaction effect accounted for 1.2% of the sample variance in the outcome variable. We consider the interpretation of b_3 shortly.

The regression coefficient associated with the $Time_c$ predictor, 0.247, is the effect of time spent together on adolescent relationship satisfaction when $Gender = 0$. Because a score of 0 on the gender variable corresponds to females, the coefficient is the effect of time spent together on adolescent relationship satisfaction for females: For every additional hour that a mother spends with her daughter, adolescent relationship satisfaction is predicted to increase 0.247 scale units on the satisfaction measure. The estimated standard error for this coefficient (taken directly from the computer output) was 0.070 and the 95% confidence interval was 0.108 to 0.386. The t test for statistical significance was $t = 3.53$, $p < .01$.

The above yields information about the effects of time spent together on satisfaction for females, but we also would like the information for males. A simple way of obtaining this information is to rescore the dummy variable so that females = 1 and males = 0, then multiply this new value by the centered time variable and rerun the analysis using these newly created variables. The resulting regression equation is

$$Y = 2.375 + -0.013 Time_c + 5.270 Gender_r$$
$$+ 0.260(Time_c)(Gender_r) + e$$

Note that the coefficient for b_3 is unchanged in magnitude but is opposite in sign. This is a result of the reverse scoring, and the reason for it will be made explicit shortly. The coefficient for *Time*, –0.013, is the effect of time spent together on adolescent satisfaction when the reversed-scored gender variable equals 0. Because a score of 0 on this variable corresponds to males, it is the effect of time spent together on adolescent relationship satisfaction for males: For every additional hour that a mother spends with her son, adolescent relationship satisfaction is predicted to change trivially, namely, 0.013 scale units on the satisfaction measure. The estimated standard error for this coefficient (taken from the computer output) was 0.089 and the 95% confidence interval was 0.189 to 0.163. The t test for statistical significance was nonsignificant ($t = 0.15$, $p > .88$).

The analyses of these simple effects reveal that the slope of adolescent satisfaction regressed onto time spent together for males is –0.013, whereas for females, it is 0.247. If we calculate the difference between these two slopes, we obtain $-0.013 - 0.247 = -0.260$. Note that this is the value of the coefficient for the product term in the original analysis, that is, $b_3 = -0.260$. *In a traditional product-term analysis where the product term involves a*

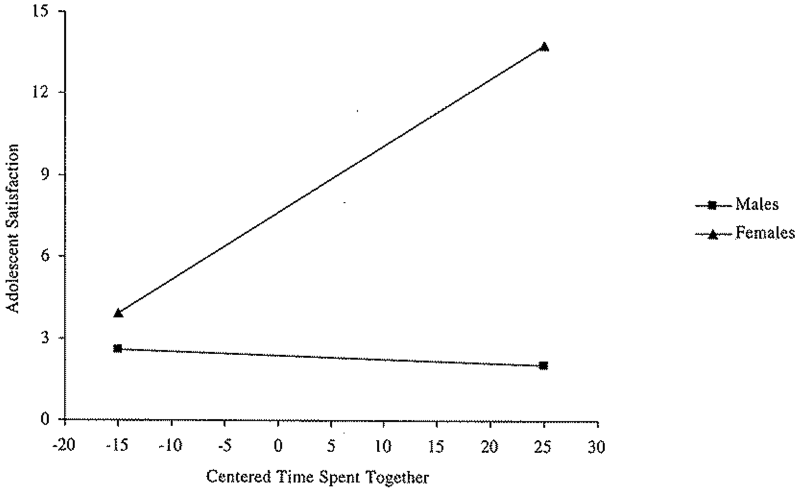

Figure 2.2 Regression Lines for Adolescent Relationship Satisfaction as a Function of Time Spent Together for Males and Females

continuous variable (X) and a dummy variable (Z), the coefficient for the product term will equal the slope difference for the regression of Y *onto* X *between the group scored 1 on the dummy variable minus the reference group on the dummy variable.* The significance test associated with b_3 is a test of slope differences and represents a single-degree-of-freedom interaction contrast. The reason the b_3 coefficient reversed itself in sign when gender was recoded was because in the first analysis, b_3 corresponded to $-0.013 - 0.247 = -0.260$, whereas in the second analysis, b_3 corresponded to $0.247 - (-0.013) = 0.260$.

In sum, we used product-term regression to calculate the slope of Y on X for females and then, through rescoring of the dummy variable, the slope of Y on X for males. We characterized the significance tests and confidence intervals for each of these slopes. We also formally tested the difference between the slopes through the regression coefficient associated with the product term.

Graphical presentations of the interaction can be displayed by plotting the regression lines for Y on X for each group defined by the moderator variable on the same graph. This has been done in Figure 2.2. An interaction effect is indicated by nonparallel lines. The intercept for the linear equation of a given group (e.g., males) is the intercept from the full equation where that group was the reference group on Z. For females, it is 7.645, and for males, it is 2.375. (Note that the intercepts in Figure 2.2 occur at the value of $X = 0$, in the middle of the graph, because the Y axis has been "moved" to the left for aesthetic purposes.)

In the applied research literature, scientists sometimes approach such group comparisons by calculating separate regression equations for each group and testing for the significance of the slope of Y on X in each group separately. A group difference is said to occur if the slope is statistically significant in one group but not in the other. This usually is poor analytic practice. Most important, the analysis does not result in a formal test of the difference in slopes between the two groups and such a test is necessary if one is going to speak of group differences. A slope may be statistically significant in one group but not in the other and still yield a nonsignificant slope difference when the groups are formally compared. Or both slopes might be statistically significant and the test of slope difference might be statistically nonsignificant. Or the two slopes could both be statistically nonsignificant but the slope difference could be statistically significant. One needs to conduct a formal comparison of slope differences to speak of group differences in slopes, and the product-term strategy does so. A second difficulty with the separate group strategy is that it ignores available information when estimating the variance of the residuals, that is, when estimating error terms for the significance tests. In the product-term analysis, the pooled estimate of the residual variance at given values of the predictor variables is based on data from all groups involved in the analysis. In the separate group strategy, it is based on only a single group.

A Continuous Moderator Variable

Suppose that the investigator conceptualized the analysis differently, reversing the roles of the focal independent variable and the moderator variable. Now the concern is with gender differences in adolescent relationship satisfaction and whether these differences vary as a function of how much time the mother spends with her child. This analysis uses the same regression equation as above, where we regressed adolescent satisfaction onto the centered time variable, gender, and the product of the two. We just focus on different aspects of the model when characterizing the interaction and conceptualize things a bit differently. Recall that the regression equation was

$$Y = 7.645 + 0.247 Time_c + -5.270 Gender$$
$$+ -0.260(Time_c)(Gender) + e$$

As before, the b_3 coefficient contains information about the interaction, that is, how the gender difference varies as a function of time spent together. We return to its interpretation shortly. First, examine the coefficient for gender. Because gender is a dummy variable, the coefficient associated with it represents a mean difference, namely, it is the mean relationship satisfaction for males (the group scored 1 on the dummy variable) minus the mean relationship satisfaction for females (the reference group) when $Time_c = 0$. Because $Time_c$ is the mean-centered *Time* variable, a score of 0 on $Time_c$ corresponds to the sample mean on the original *Time* variable (which was 24.06 hr). Our estimate of the mean difference in relationship satisfaction between males and females (−5.270) is for the case where the amount of time that mothers spend with their children equals 24.06 hr. This gender difference is statistically significant based on the t test of b_2 yielded by the computer output ($t = 10.28$, $p < .01$). The estimated standard error of the difference is 0.513 and the 95% confidence interval is −6.288 to −4.253. The intercept term, 7.645, is the estimated mean value of Y when all predictors equal 0. If $Time_c = 0$ and $Gender = 0$, all terms in the equation become 0, so 7.645 is the mean satisfaction for females when *Time* equals its sample mean. If the difference between the male mean and the female mean is −5.270 and if the female mean is 7.645, then simple algebra allows us to calculate the value of the male mean. It is the difference plus the female mean, or $-5.270 + 7.645 = 2.375$. Note that 2.375 (the male mean) minus 7.645 (the female mean) equals the mean difference, −5.270.

What happens to the estimated value of the gender difference in relationship satisfaction at a different value of time spent together? From the above analysis, we know that the gender difference is 5.270 when the time spent together is 24.06 hr. What is it at some other value? For the sake of pedagogy, let us calculate the gender difference at a value of 25.06 hr; that is, let us increase the hours spent together from the previous analysis by one unit. We can do so using our transformation strategy. We first transform the *Time* variable so that a score of 0 on the transformed variable corresponds to a score of 25.06 on the original time variable. This is accomplished by subtracting the value of 25.06 from *Time*. The transformed time variable, $Time_t$, is then multiplied by the dummy variable for gender, and the three terms are entered into a regression equation. The results are

$$Y = 7.892 + 0.247 Time_t + -5.530 Gender$$
$$+ -0.260(Time_t)(Gender) + e$$

The coefficient for gender is −5.530, which reflects the gender difference in mean relationship satisfaction when $Time_t = 0$. This means that when

Time = 25.06 hr, the gender difference is −5.530. Compare this with the previous analysis: When the time spent together was 24.06 hr, the gender difference was −5.270. When we increased the time spent together by one unit to 25.06 hr, the difference changed by −0.260 units to a value of −5.531. Now examine the value of b_3. It equals −0.260, which is the amount that the mean difference changed when we increased the moderator variable by one unit. *In a traditional product-term analysis where the product term involves a continuous variable (Z) and a dummy variable (X), the coefficient for the product term reflects how much the mean difference between the group scored 1 on the dummy variable minus the reference group is predicted to change given a one-unit increase in the continuous variable.* In this second analysis, the intercept term, 7.892, is the mean satisfaction for females when *Time* = 25.06. We can calculate the mean satisfaction for males in this case using algebra. It is the mean difference plus the female mean, −5.531 + 7.892 = 2.361.

Although the regression equation contains all of the relevant information one desires, some investigators will report not only the regression equation and its associated statistics but also the estimated mean difference for the focal independent variable at two or three selected values of the continuous moderator variable. This helps give the reader an intuitive feel for the interaction effect. For example, we used the transformation strategy to derive the following illustrative statistics:

Time Together	*Male Mean*	*Female Mean*	*Mean Difference*	*95% CI for Difference*	*t*	*p Value*
20 hr per week	2.43	6.64	−4.21	−5.54 to −2.98	6.75	< .001
25 hr per week	2.36	7.88	−5.52	−6.59 to −4.44	10.18	< .001
30 hr per week	2.30	9.11	−6.82	−8.63 to −5.00	7.46	< .001

At 20 hr per week spent together, the gender difference in relationship satisfaction was −4.21 (with females being 4.21 units more satisfied than males). When the amount of time spent together increased to 25 hr per week, the gender difference grew to −5.52 units. At 30 hr per week, it increased even more, to −6.82. These illustrative statistics provide an intuitive sense of the interaction, as one sees varying mean differences depending on the value of the moderator variable. However, it should be kept in mind that these simple effects are not the crux of the test of the interaction. The formal test of the interaction is captured in b_3.

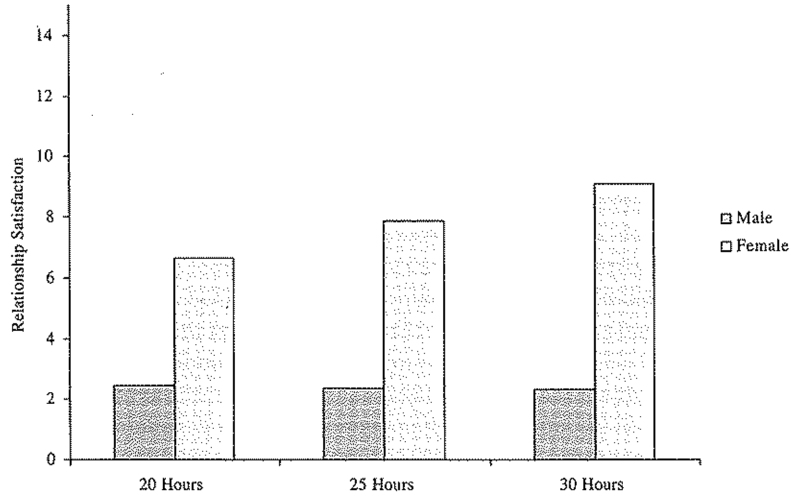

Figure 2.3 Predicted Mean Relationship Satisfaction as a Function of Gender at Selected Values of Time Spent Together

Graphical Displays. The results of these analyses can be presented graphically using bar graphs. This is illustrated in Figure 2.3. If there was no interaction, the difference in bar height between males and females would be the same at each level of time spent together. Lack of uniformity of such differences suggests an interaction. It can be seen in the graph that as the amount of time spent with their children increases, the gender difference in relationship satisfaction widens. As an alternative graphical presentation, one can plot the separate regression lines for Y on Z for the different groups defined by the focal independent variable. This was done in Figure 2.2 in the previous section. The distance between the two regression lines at any given point on the horizontal axis for the moderator variable reflects the mean difference in Y at that point. This graphical display contains more information than that of Figure 2.3, but some researchers find the former more intuitively appealing.

More Than Two Groups for the Qualitative Variable

Some applications involving qualitative and continuous predictors include qualitative variables with more than two levels. The principles described above generalize directly to these situations. Consider the case where the outcome variable is income (Y) and the two predictor variables are the number of years of education and ethnicity. Suppose that ethnicity has three levels, African American, European American, and Latino. It is represented by two dummy variables. Let the first dummy variable, D_{AA}, be defined as one where all African Americans receive a 1 and everyone else is assigned a 0. The second dummy variable, D_L, assigns all Latinos a 1 and everyone else a 0. European Americans are the reference group. To represent the interaction between the two predictor variables, all possible pairs of product terms are created between the variables representing one predictor and the variables representing the other predictor. Letting *ED* signify the measure of education, we form the product terms $(ED)(D_{AA})$ and $(ED)(D_L)$. The overall test of the omnibus interaction effect between education and ethnicity requires the use of the hierarchical regression test of Equation 1.1. The squared multiple correlation for the main-effect-only model (the model that includes only ED, D_{AA}, and D_L) is compared against the interaction model that includes both the main effects and all of the product terms. It is not sufficient to examine the significance tests for the individual b coefficients associated with the product terms to make a statement about the statistical significance of the omnibus interaction effect. This is because the omnibus interaction effect has more than a single degree of freedom and must be evaluated using the hierarchical strategy.

How one orients to the regression equation itself depends on which variable is chosen as the moderator variable. Consider first the case where ethnicity is the moderator variable (i.e., the moderator variable is the qualitative variable). In this scenario, the researcher is interested in examining the impact that education has on income and how this varies as a function of ethnicity. The regression equation has the general form

$$Y = a + b_1ED + b_2D_{AA} + b_3D_L + b_4ED(D_{AA}) + b_5ED(D_L) + e \qquad [2.12]$$

As discussed previously, the regression coefficients for the product terms reflect differences in slopes. The coefficient b_4 focuses on the slope of Y on ED and estimates the difference between this slope for the group scored 1 on D_{AA} minus the corresponding slope for the reference group. In other words, it reflects the slope difference between African Americans and European Americans. The coefficient b_5 corresponds to the same slope difference but for Latinos versus European Americans. The significance tests, estimated standard errors, and confidence intervals for these coefficients provide the additional information that usually is of interest for these single-degree-of-freedom interaction contrasts.

What if, in addition to the two contrasts isolated by b_4 and b_5, we also want to evaluate the contrast that compares the slope of Y on ED for African Americans versus Latinos? This contrast and its relevant significance tests can be calculated by creating a new set of dummy variables and product terms that change the reference group for ethnicity to be either African Americans or Latinos rather than European Americans, then rerunning the regression analysis and examining the product-term coefficient in the new equation that isolates the comparison. We discuss in Chapter 4 the issue of conducting multiple contrasts and adjustments for inflated error rates across contrasts.

The coefficient for ED in Equation 2.12, b_1, is a simple effect. It is the effect of education on income when all the ethnicity dummy variables equal 0. Scores of $D_{AA} = 0$ and $D_L = 0$ map onto the reference group, so b_1 is the effect of education on income for European Americans. The estimated standard error, confidence interval, and significance test for this effect can be taken from the computer output for b_1. If one desires to isolate the corresponding simple effect and significance test for Latinos, then simply redefine a new set of dummy variables and product terms so that Latinos are the reference group and rerun the analysis on the computer. The process is similar for African Americans.

Next, consider the case where education is the moderator variable. Now the interest is in ethnic differences in income and how these differences might vary as a function of education. The same regression equation is used, but researchers typically will mean center education (or center it around a meaningful value) prior to using it in the analysis. We assume that mean centering has been invoked in the following discussion.

Consider first the coefficient for D_{AA}, which is b_2. This coefficient is the estimated mean income difference between the group scored 1 on this dummy variable and the reference group when $ED_c = 0$. It is the estimated mean income difference between African Americans and European Americans when education equals the value of its sample mean. The coefficient b_4 for the product term of D_{AA} and ED_c indicates how this mean difference changes given a one-unit increase in education. The coefficient for D_L is b_3. This coefficient is the estimated mean income difference between Latinos and European Americans when education equals the value of its sample mean. The coefficient b_5 for the product term of D_L and ED_c indicates how this mean difference changes given a one-unit increase in education. If one is interested in the analogous parameters and significance tests for the mean difference between African Americans and Latinos, then

TABLE 2.4
Two Interaction Forms

Value of Z	Example A Mean Difference Between Two Groups	Example B Mean Difference Between Two Groups
1	2.0	2.0
2	4.0	4.0
3	6.0	6.0
4	8.0	8.0
5	10.0	10.0
6	12.0	10.0
7	14.0	10.0

redefine the dummy variables so that either African Americans or Latinos are the reference group and rerun the analysis on the computer. Then isolate the relevant coefficients using the principles discussed earlier.

The intercept in the regression equation is the estimated mean income for the reference group (European Americans) when $ED_c = 0$.

Form of the Interaction

As noted earlier, when both variables involved in the product term are continuous, the traditional interaction model evaluates an interaction of a particular form, namely, a bilinear interaction. When one of the variables in the product term is qualitative and the other is continuous, the traditional interaction model also evaluates a specific type of interaction. The model can be stated from two different vantage points, although both perspectives reflect the same underlying model. One vantage point is when the moderator variable is the qualitative variable. In this case, the assumption is that the relationship between *Y* and *X* (the outcome variable and the continuous focal independent variable) is linear in form and this is true at each level of the moderator variable. If the relationship between *Y* and *X* is nonlinear for at least one level of the moderator variable, then the traditional product-term model is misspecified and an alternative modeling approach is necessary (discussed in Chapter 4).

Another vantage point is when the moderator variable is the continuous variable. In this case, the assumption is that the orderly changes in the group mean differences as a function of *Z* are a linear function of *Z*. Consider the two examples in Table 2.4. Example A illustrates a case where the mean difference between two groups changes linearly with *Z*: For every one unit that Z increases, the mean difference between the groups becomes two units larger. Example B reflects nonlinear change and should not be modeled with the traditional product-term approach: At low levels of Z, for every one unit that Z increases, the mean difference between the groups becomes two units larger, but this eventually asymptotes and levels off.

Summary

In sum, researchers often are interested in exploring interaction models that focus on two-way interactions. The variables involved in the interaction might both be continuous in nature, or one might be continuous and the other qualitative. In either case, the researcher declares one of the variables the focal independent variable and the other the moderator variable. Appropriate product terms are calculated and then the interaction model is estimated including all of the "main-effect" terms as well as all of the product terms. The significance test of the omnibus interaction is reflected in the regression coefficient associated with the product term when the omnibus interaction has a single degree of freedom. When the omnibus interaction has more than a single degree of freedom, the hierarchical test that compares the main-effect model with the interaction model needs to be applied to evaluate the statistical significance of the overall interaction.

The strength of the interaction effect in unstandardized terms is indicated in the value of the regression coefficients reflecting the single-degree-of-freedom interaction contrasts. The strength of the effect in standardized terms is indexed by the squared semipart correlation for the interaction term(s).

Interpretation of the interaction typically focuses on the regression coefficients associated with the focal independent variable and the regression coefficients associated with the product terms. Depending on which variable is the moderator variable, different centering and recoding strategies can be used to help the reader appreciate the nature of the interaction.

3. THREE-WAY INTERACTIONS

This chapter extends the principles of the previous chapter to the analysis of three-way interactions. We consider first the case of all continuous predictors and then consider the case of a mixture of qualitative and continuous predictors.

Three Continuous Predictors

Consider a study where a researcher in behavioral medicine studied a parent's intention to vaccinate his or her child to prevent a certain form of hepatitis. The intention measure ranges from 0 to 100, with higher scores indicating greater intent to obtain the vaccination. The researcher studied three variables as potential predictors of intent to vaccinate. The first variable was the perceived likelihood that the child would contract the disease if the parent did not obtain the vaccination. This variable is called perceived susceptibility to the disease and was measured on a 0 to 100 scale, with higher numbers indicating higher levels of perceived susceptibility. The second variable was how serious the parent thought it would be if the child did in fact contract hepatitis. This variable is called perceived severity and was measured on a 0 to 100 scale, with higher numbers indicating higher levels of perceived severity. The third variable was the perceived likelihood that the vaccination would prevent the disease. This variable is called perceived efficacy and was measured on a 0 to 100 scale, with higher numbers indicating higher levels of perceived efficacy. The intention to vaccinate was thought to be an interactive function of these three variables. The sample size for the study was 200 parents.

To analyze a three-way interaction, it is helpful to first specify a focal independent variable and the moderator variables and to explicate the underlying logic of the interaction. For three-way interactions, we need to make distinctions among the moderator variables because there are two of them. Suppose that the focal independent variable chosen by the investigator was the perceived susceptibility variable. The researcher hypothesized that the intent to vaccinate would vary as a function of perceived susceptibility to the disease. Specifically, he hypothesized that intentions to vaccinate would be higher as the perceived susceptibility to the disease increased. He further reasoned that the effect of perceived susceptibility to the disease on intent to vaccinate would be moderated by the perceived severity of the disease. If a parent thought the disease was not very serious, then it would not matter how susceptible the child was to it. The parent would not be motivated to seek a vaccination for a disease that the parent deemed inconsequential. As the perceived severity of the disease increased, however, then it was felt that variations in perceived susceptibility would impact intent to vaccinate. In this context, perceived severity is called a *first-order moderator variable* because it is thought to directly moderate the effect of perceived susceptibility on intent to vaccinate. The researcher further hypothesized that the moderating effects of perceived severity would differ depending on the perceived efficacy of the vaccination. If parents felt that the vaccination was not efficacious, then they would not seek to vaccinate their child no matter what the susceptibility to the disease or how severe the disease is. However, as perceived efficacy of the vaccination increases, then the interactive dynamics between severity and susceptibility would manifest themselves. The perceived efficacy of the vaccination is a *second-order moderator variable* because it moderates the impact of the first-order moderator on the relationship between the focal independent variable and the dependent variable. Of course, it is not necessary to conceptualize three-way interactions in these terms. However, we have found that invoking the concepts of first-order and second-order moderators is useful for organizing one's thinking about three-way interactions. In addition, when investigators describe the results of three-way interactions, they almost always adopt such an orientation (albeit sometimes implicitly) in order to make sense of the complex relationships involved.

To analyze the traditional interaction model, we need to form product terms that, when added to the main-effect model, reflect two-way interactions and the three-way interaction. The main-effect model is

$$Y = a + b_1X + b_2Z + b_3Q + e$$

where X is perceived susceptibility, Z is perceived severity, and Q is perceived efficacy. The three-way interaction model adds to this all possible pairwise product terms among the three predictors (XZ, XQ, and ZQ) as well as a product term for all three predictors (XQZ). This results in the following model:

$$\begin{aligned} Y = a + b_1X + b_2Z + b_3Q + b_4XZ + b_5XQ \\ + b_6QZ + b_7XZQ + e \end{aligned} \qquad [3.1]$$

The significance test of the three-way interaction is the significance test of b_7. The strength of the three-way interaction in standardized terms is found by subtracting from the squared multiple correlation of Equation 3.1 the value of the squared multiple correlation for a model based on Equation 3.1 that omits XZQ. The interpretation of the lower-order coefficients always are conditionalized on the higher-order product terms, with the conditionalization being that the other variables in the higher-order product terms equal 0. For example, the coefficient b_4 reflects the effect of the two-way interaction between X and Z on Y when $Q = 0$. The coefficient b_1 reflects the effect of X on Y when $Z = 0$ and $Q = 0$. The distinction between the focal independent variable, the first-order moderator variable, and the second-order moderator variables guides us in how we orient to the coefficients in the equation.

Returning to our example, we begin by mean centering all of the predictors and then forming the relevant product terms. The resulting regression output from SPSS is shown in Table 3.1. The coefficient for the three-way product term, 0.0008137, is statistically significant ($t = 4.715$, $p < .01$, 95% CI = 0.000473 to 0.001154), suggesting the presence of a three-way interaction. The squared multiple correlation for the model that omits the three-way product term is .400, whereas the full model that includes all the product terms including the three-way term yields a squared multiple correlation of .462. The difference between these squared multiple correlations is .062, indicating that the three-way interaction accounts for 6.2% of the variance in the intent to vaccinate.

Based on the principles presented earlier, the coefficient for susceptibility (X), 0.219, is the effect of perceived susceptibility on intent to vaccinate when perceived severity (Z) and perceived efficacy (Q) equal 0 (i.e., when perceived severity and perceived efficacy are "average" or "medium" based on the fact that we mean centered them). The coefficient of 0.219 indicates that for every one unit that perceived susceptibility increases, the intent to vaccinate is predicted to increase by 0.219 units ($t = 5.062$, $p < .01$, 95% CI = 0.134 to 0.304). The coefficient for XZ reflects the two-way interaction between perceived susceptibility and the first-order moderator variable, perceived severity, when the second-order moderator variable, perceived efficacy, equals 0 (i.e., when perceived efficacy is "medium" or "average"). The coefficient was 0.01227 ($t = 4.61$, $p < .01$, 95% CI = 0.007 to 0.018). This is the amount by which the slope of intent to vaccinate on perceived susceptibility (i.e., Y on X) is predicted to change given a one-unit increase in perceived susceptibility when perceived efficacy is "average." For every one unit that perceived severity increases, the slope of intent to vaccinate on perceived susceptibility increases by 0.01227 units, holding perceived efficacy constant at its sample mean. This coefficient is interpreted just like a two-way interaction as described in Chapter 2, but it is conditionalized on $Q = 0$ (in this case, when the centered perceived efficacy equals 0).

To gain an appreciation for the meaning of the coefficient for the three-way product term, suppose that we recalculate the coefficient for XZ for the case where perceived efficacy (Q) is one unit above its sample mean rather than at its sample mean. This can be accomplished by subtracting a constant of 45.79 from the perceived efficacy score rather than the sample mean of 44.79, thus defining the zero point on a transformed efficacy scale as 45.79 on the original efficacy scale. We then recalculate the product terms and rerun the regression analysis using the transformed scores. As discussed in

TABLE 3.1
SPSS Output for Regression Analysis of Three-Way Interaction

Model	*R*	*R^2*	*Adjusted R^2*	*SE of the Estimate*
1	.680	.462	.443	9.85080

Model	*Unstandardized Coefficients*		*Standardized Coefficients*	*t*	*Sig.*	*95% Confidence Interval for B*	
	B	*SE*	*β*			*Lower Bound*	*Upper Bound*
(Constant)	33.332	.697		47.849	.000	31.958	34.706
Susceptibility (*X*)	.219	.043	.270	5.062	.000	.134	.304
Severity (*Z*)	.203	.044	.245	4.629	.000	.116	.289
Efficacy (*Q*)	.248	.044	.299	5.635	.000	.161	.334
XZ	.01227	.003	.246	4.606	.000	.007	.018
XQ	.01254	.003	.243	4.519	.000	.007	.018
QZ	.01383	.003	.263	4.951	.000	.008	.019
XQZ	.0008137	.000	.254	4.715	.000	.000	.001

Chapter 2, the squared multiple correlation and the value of b_7 are invariant to this transformation. The coefficient for the *XZ* product term in this new analysis is 0.0130837. We can index the change in the two-way interaction parameter that occurred from the previous analysis by calculating the difference between the two b_4 coefficients. This is $0.0130837 - 0.01227 = 0.0008137$. Examine the coefficient for the three-way product term in Table 3.1. It equals 0.0008137, the amount by which the two-way interaction coefficient changed given a one-unit increase in perceived efficacy. *For a three-way interaction model with three continuous predictors,* X, Z, *and* Q, *and the various product terms between them, let* X *be the focal independent variable,* Z *be the first-order moderator variable, and* Q *be the second-order moderator variable. The coefficient for the three-way product term is the predicted change in the two-way interaction parameter for* X *and* Z *given a one-unit increase in* Q. If we reran the analysis centering perceived efficacy two units above its sample mean, then the coefficient for *XZ* would be $0.01227 + 0.0008137 + 0.0008137 = 0.0138974$.

Although all the requisite information for interpretation of the three-way interaction is contained in the regression equation, most readers have a difficult time gaining an intuitive sense of the three-way interaction from inspection of the equation. As in Chapter 2, we have found it helpful to provide estimated values of the slope of *Y* on *X* at different combinations of *Z* and *Q* in order to assist the reader in thinking about the interaction. This has been done in Table 3.2. In this table, we created four scenarios under which to describe the slope of *Y* on *X:* (1) a low value on *Z* and a low value on *Q*, (2) a low value on *Z* and a high value on *Q*, (3) a high value on *Z* and a low value on *Q*, and (4) a high value on *Z* and a high value on *Q*. The "low" and "high" values were defined as 1 *SD* below the sample mean and 1 *SD* above the sample mean of *Q* and *Z*, respectively. The slope of *Y* on *X* for each of these scenarios is presented as entries in a 2×2 factorial table where the low and high values of the first-order moderator variable are rows and the low and high values of the second-order moderator variable are columns. The values of these coefficients, their significance tests, and their confidence intervals all were calculated by computer using the transformation strategy to isolate the relevant coefficient for susceptibility under the different scenarios. This table shows that when perceived efficacy is relatively low, the difference in effects of perceived susceptibility on intent to vaccinate as a function of perceived severity are negligible and small compared with when perceived efficacy is relatively high. Indeed, it is only when both perceived severity and perceived efficacy are relatively high that one begins to see a meaningful impact of perceived susceptibility on the intent to vaccinate.

TABLE 3.2
Slopes for Intent to Vaccinate on Perceived Susceptibility as a Function of Perceived Severity and Perceived Efficacy

	Low Efficacy			*High Efficacy*		
	Slope	*95% CI*	*t*	*Slope*	*95% CI*	*t*
Low severity	0.003	−0.128 to 0.188	0.376	0.009	−0.183 to 0.200	0.089
High severity	0.016	−0.145 to 0.178	0.197	0.820	0.648 to 0.993	9.39*

NOTE: SE = estimated standard error; CI = confidence interval.
*$p < .01$.

These illustrative statistical portrayals also can be presented graphically using a side-by-side plot. The separate regression lines as a function of the two values of Z are plotted on one graph when Q is "low" and then again on a separate graph when Q is "high." The two graphs are then presented "side by side." An example of such a plot is presented in Figure 3.1. If there is no three-way interaction, then the divergence of the two slopes from parallelism in one graph should be the same as that in the other graph. This is clearly not the case for our data.

To construct the above plot, you need the intercept for the regression of Y on X at selected values of Z and Q. The intercept will be the value of the intercept term in the full interaction regression equation that uses the transformed values of Z and Q to isolate the relevant slope of interest.

Qualitative and Continuous Predictors

Consider a study where a researcher is interested in prejudice in the judicial system. He designed a study where participants read a scenario describing a court case against a defendant and then rated, on a 100-point scale, the likelihood that the defendant was guilty of the crime. The outcome variable ranged from 0 to 100, with higher scores indicating a higher probability of guilt or a higher tendency to attribute guilt. All participants read the identical scenario except that half learned that the accused was African American while the other half learned that the accused was European American. Half of the study participants were themselves African American and half were European American. The result was a 2×2 factorial design that crossed the ethnicity of the accused with the ethnicity of the "juror" (i.e., the study

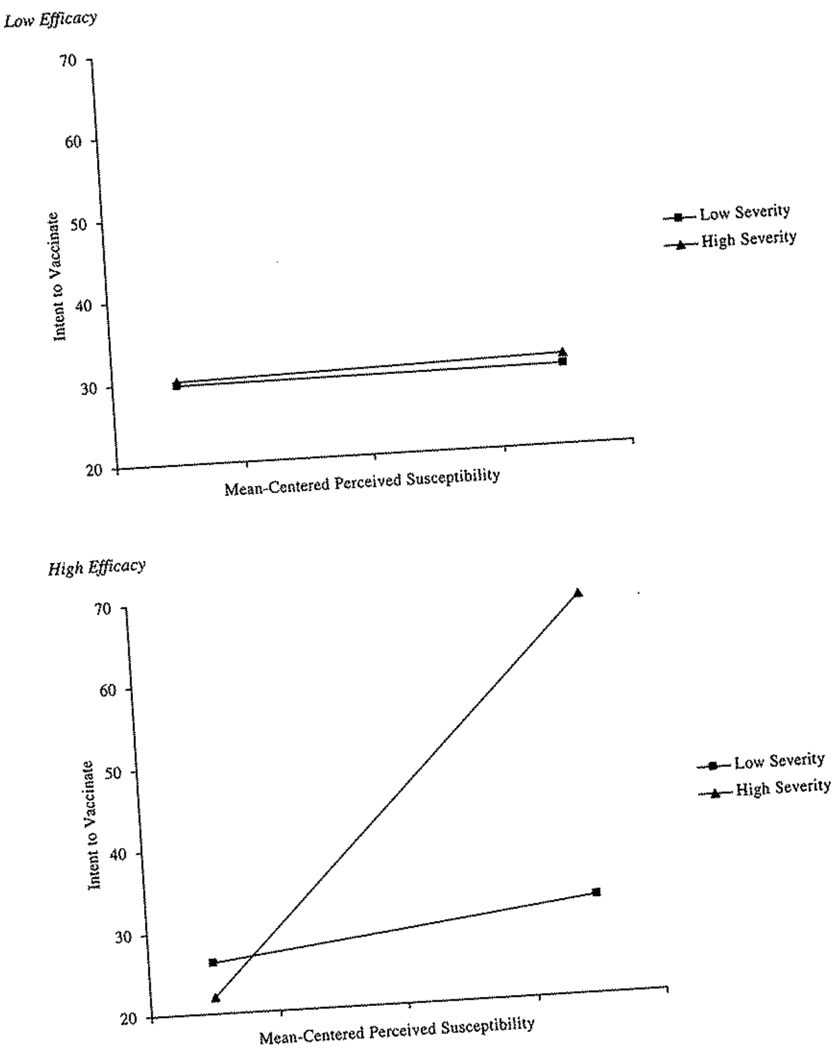

Figure 3.1 Side-by-Side Plot for the Regression Lines of Intent to Vaccinate on Perceived Susceptibility as a Function of Perceived Severity and Perceived Efficacy

participant). The researcher also measured a third variable, namely, how liberal or conservative the research participant was. The relevant scale ranged from 0 to 100, with higher scores indicating higher levels of liberalness. Ethnicity of the accused was represented by a single dummy variable, *AC*, with European Americans representing the reference group. Ethnicity of the juror also was represented by a single dummy variable, *JU*, with European Americans as the reference group. The liberalness scale, *L*, was mean centered and the relevant product terms formed (hereafter, *L* refers to the centered liberalness-conservativeness scale). Table 3.3 presents the output from an SPSS regression program.

The coefficient for the three-way product term, 1.007, is statistically significant ($t = 7.31$, $p < .01$, 95% CI = 0.735 to 1.279), suggesting the presence of a three-way interaction. The squared multiple correlation for the model that omits $AC \times JU \times L$ is .595, whereas the full model yields a squared multiple correlation of .782. The difference between these squared multiple correlations is .192, indicating that the three-way interaction accounts for 19.2% of the variance in the guilt attributions. We discuss the coefficients in the model from two vantage points, one where the continuous variable is the focal independent variable and one where one of the qualitative variables is the focal independent variable.

A Continuous Focal Independent Variable

In this framework, the researcher is interested in the effects of liberalness-conservativeness on guilt attributions. She posits that more liberal jurors will be less likely to ascribe guilt to the accused than more conservative jurors. However, the effect of liberalness-conservativeness on guilt attributions is hypothesized to be moderated by the ethnicity of the accused: When the accused is African American, liberalness-conservativeness will be a stronger predictor of guilt attributions than when the accused is European American. Thus, the ethnicity of the accused is a first-order moderator. In accord with a three-way interaction, the way in which the ethnicity of the defendant moderates the effect of liberalness-conservativeness on guilt attributions is thought to depend on the ethnicity of the juror. It is predicted that the aforementioned dynamics for the two-way interaction between ethnicity of the defendant and liberalness-conservativeness will be more pronounced for European American jurors than for African American jurors. This is because prejudice and race are thought in judicial cases to be more salient to European American jurors than African American jurors. Ethnicity of the juror is thus a second-order moderator variable.

The coefficient for liberalness-conservativeness in Table 3.3, −0.576, reflects the effects of liberalness-conservativeness on guilt attributions

TABLE 3.3
SPSS Output for Juror Example

Model	*R*	R^2	*Adjusted* R^2	*SE of the Estimate*
1	.884	.782	.775	8.58238

Model	*Unstandardized Coefficients*		*t*	*Sig.*	*95% Confidence Interval for B*	
	B	*SE*			*Lower Bound*	*Upper Bound*
(Constant)	39.006	1.111	35.102	.000	36.817	41.196
Accused Ethnicity (*AC*)	17.235	1.570	10.980	.000	14.142	20.328
Juror Ethnicity (*JU*)	1.228	1.570	.782	.435	−1.865	4.320
Liberalness (*L*)	−.576	.068	−8.473	.000	−.710	−.442
AC * *JU*	−17.241	2.218	−7.772	.000	−21.611	−12.870
AC * *L*	−.952	.098	−9.717	.000	−1.145	−.759
JU * *L*	−.033	.098	−.340	.734	−.226	.160
AC * *JU* * *L*	1.007	.138	7.307	.000	.735	1.279

TABLE 3.4
Slopes for Guilt Attributions Regressed Onto Liberalness

	Slope	*SE*	*95% CI*	*t*	*p Value*
AA juror, AA defendant	−0.554	0.066	−0.685 to −0.424	8.35	< .001
AA juror, EA defendant	−0.609	0.071	−0.748 to −0.470	8.64	< .001
EA juror, AA defendant	−1.528	0.071	−1.667 to −1.389	21.65	< .001
EA juror, EA defendant	−0.576	0.068	−0.710 to −0.442	8.473	< .001

NOTE: AA = African American; EA = European American; SE = estimated standard error; CI = confidence interval.

when $AC = 0$ and when $JU = 0$. This corresponds to the case of the European American defendant as judged by the European American jurors. It is a simple effect. We can make explicit the meaning of the other coefficients in the equation if we first isolate the corresponding simple effects for the remaining three conditions. This is accomplished by recomputing a new set of dummy variables for AC and JU that redefine the reference groups so that the coefficient associated with the liberalness-conservativeness effect reflects the simple effect of interest. For example, if we define the reference group on AC as the European American defendant and the reference group on JU as African American jurors, recalculate all the product terms, and then rerun the regression analysis, the coefficient associated with L will reflect the effect of liberalness-conservativeness on guilt attributions for the case of a European American defendant as rated by African American jurors. Table 3.4 presents the relevant slopes for the four conditions as derived from the various computer analyses.

Examine in Table 3.3 the coefficient for the product term of the focal independent variable and the first-order moderator variable, $AC \times L$. This coefficient reflects a two-way interaction between a categorical variable and a continuous variable, so it reflects a difference in slopes. Specifically, it is the slope of guilt attributions regressed onto liberalness-conservativeness for an African American defendant minus the same slope for a European American defendant when $JU = 0$. The value of 0 on JU corresponds to European American jurors, so it is the slope difference focusing on only European Americans jurors. The coefficient equals −0.952. Examine Table 3.4 and focus on only the European American jurors. Note that the slope of guilt attribution on liberalness-conservativeness for the African American defendant is −1.528 and for the European American defendant is −0.576. The difference between these two slopes is $(-1.528) - (-.576) = -0.952$, the value of the two-way interaction parameter. From the output in Table 3.3, this difference in slopes is statistically significant ($t = 9.72$, $p < .01$, 95% CI = −1.145 to −0.759). *For an interactive model with a continuous predictor,* X, *and two qualitative predictors,* Z *and* Q, *and the corresponding product terms between them, let* X *be the focal independent variable,* Z *be the first-order moderator variable, and* Q *be the second-order moderator variable. For dummy coding on the qualitative predictors, the regression coefficient for the* XZ *product term is a difference in slopes. The difference focuses on the slope of* Y *on* X *for the group scored 1 on the dummy variable for* Z *minus the corresponding slope for the reference group on* Z, *all with respect to the reference group on* Q.

If the slope difference between the African American defendant and the European American defendant is −0.952 for European American jurors, what is it for African American jurors? The value of this difference, its estimated standard error, confidence interval, and significance test can be computed by redefining the *JU* dummy variable so that African American jurors are the reference group, recalculating the product terms, rerunning the regression analysis, and then examining the coefficient for the $AC \times L$ product term in the new analysis. The coefficient in the new analysis is 0.055. Confirm in Table 3.4 that the relevant slopes are −0.554 and −0.609 and that their difference is (−0.554) − (−0.609) = 0.055. This difference is not statistically significant ($t = .565$, $p < .57$, 95% CI = −0.136 to 0.246).

To summarize thus far, for African American jurors, the slope difference of *Y* on *X* for the African American defendant minus the European American defendant is 0.055. This is the estimated two-way interaction parameter for African American jurors. The corresponding slope difference for European American jurors is −0.952. It represents the estimated two-way interaction parameter for European American jurors. If there is no three-way interaction, then these two-way parameter values should be the same (assuming no sampling error). The fact that they are not suggests that there may be a three-way interaction. The difference between the two-way interaction parameters is (0.055) − (−.952) = 1.007. Examine the coefficient for the three-way product term. Note that it equals 1.007. The significance test for this coefficient evaluates the probability of observing a result of 1.007 or greater assuming that the null hypothesis of a zero difference in the two-way interaction parameters is true. *For an interactive model with a continuous predictor,* X, *and two qualitative predictors,* Z *and* Q, *and the corresponding product terms between them, let* X *be the focal independent variable,* Z *be the first-order moderator variable, and* Q *be the second-order moderator variable. For dummy coding on the qualitative predictors, the regression coefficient for the "three-way" product term is a difference of slope differences.*

The difference focuses on the slope of Y *on* X *for the group scored 1 on the dummy variable for* Z *minus the corresponding slope for the reference group on* Z. *This slope difference for the reference group on the dummy variable for* Q *is subtracted from the corresponding slope difference for the group scored 1 on* Q.

In sum, Table 3.4 provides the slopes of Y on X in the various experimental conditions and conveys the information about the significance tests for those simple effects. A three-way interaction implies that the nature of the two-way interaction varies depending on the value of the second-order moderator variable. The patterning of slopes in Table 3.4 makes evident the source of the three-way interaction. The slope differences for European American jurors are more dramatic than those for African American jurors. This could be shown graphically in a side-by-side plot of the regression of Y on X for each of the two groups defined by Z, with a separate plot at each value of Q.

A Qualitative Focal Independent Variable

We can reanalyze the above study from the perspective of using a qualitative predictor as the focal independent variable. Here is the conceptual logic: The researcher predicts that because of negative stereotypes and prejudice, African American defendants will tend to be seen as more likely to be guilty than European American defendants. However, this tendency is hypothesized to be qualified by the ethnicity of the juror (the first-order moderator variable): Whereas higher levels of guilt will be attributed to the African American as opposed to the European American defendant, this will be evident for only European American jurors. This hypothesis derives from the assumption that European American jurors will exhibit prejudice but not African American jurors. Finally, this moderating effect of juror ethnicity will vary depending on how liberal-conservative the juror is. For conservative jurors, the above dynamics characterizing the two-way interaction will be evident. However, for liberal jurors, this will not be the case and no prejudice will be shown for either the African American or European American jurors.

The test of the three-way interaction is exactly as in the previous section, and we again use the regression equation in Table 3.3 as our reference. However, the focus is on different lower-order coefficients to explicate the three-way interaction in line with the conceptual framework stated above.

After affirming that the three-way interaction is present, we examine first the coefficient associated with the focal independent variable and the first-order moderator variable, which is the product term $AC \times JU$. The value for this coefficient is −17.241. If both variables in a product term are dummy

TABLE 3.5
Mean Guilt Attributions as a Function of Ethnicity of the Accused and Ethnicity of the Juror When Liberalness-Conservativeness Equals Its Sample Mean

	Mean	*SE*	*95% CI*
AA juror, AA defendant	40.23	1.11	38.04 to 42.41
AA juror, EA defendant	40.23	1.11	38.04 to 42.41
EA juror, AA defendant	56.24	1.11	54.06 to 58.43
EA juror, EA defendant	39.00	1.11	36.62 to 41.20

NOTE: AA = African American; EA = European American; SE = estimated standard error; 95% CI = 95% confidence interval.

variables with dummy coding, then the coefficient associated with the product term reflects information about a 2×2 table in the design. The 2×2 table to which it refers crosses the group scored 1 on the first dummy variable versus its reference group with the group scored 1 on the second dummy variable with its reference group on *JU*. This subtable is identified in Table 3.5. It will be helpful if we identify the mean guilt attributions for each cell of the 2×2 table, conditional on the centered liberalness-conservativeness score being 0. As it turns out, the intercept in the full regression equation is the mean guilt attribution when $AC = 0$, $JU = 0$, and $L = 0$. So the mean guilt attribution for a European American defendant as rated by European American jurors when liberalness-conservativeness is "average" is 39.006 (95% CI = 36.817 to 41.196). We can calculate the other three cell means either by algebra (as discussed in Chapter 4) or by recoding the dummy variables and rerunning the analyses. In the latter approach, we redefine the dummy variables so that the reference groups correspond to the cell of the 2×2 table in which we are interested. We then recalculate the product terms, rerun the regression analysis, and then note the value of the intercept in the new equation. We used this strategy to generate the four cell means and their confidence intervals in Table 3.5.

The interaction parameter for this 2×2 table, as discussed in Chapter 1, is the mean difference for the focal independent variable at the first level of the moderator variable minus the mean difference for the focal independent variable at the second level of the moderator variable. It is $(40.23 - 40.23) - (56.24 - 39.00) = -17.24$. Note that this value is identical to the regression coefficient for $AC \times JU$. The estimated standard error, confidence interval, and significance test for the coefficient all provide perspectives on the viability of the interaction in the 2×2 table. *For an interactive model* *with two qualitative predictors,* X *and* Z*, and a continuous predictor,* Q*, and the corresponding product terms between them, let* X *be the focal independent variable,* Z *be the first-order moderator variable, and* Q *be the second-order moderator variable. For dummy coding on the qualitative predictors, the regression coefficient for* XZ *is the difference between two mean differences. It is the difference between the group scored 1 on* X *and the reference group on* X *for the group scored 1 on* Z *minus the corresponding mean difference for the reference group on* Z*, all when* Q = *0.*

The coefficient for the three-way interaction term tells us how this two-way interaction parameter changes given a one-unit increase in liberalness-conservativeness. The coefficient for the three-way product term was 1.007. This means that if we increased liberalness by one unit, the aforementioned interaction parameter of −17.241 would now equal −17.241 + 1.007 = −16.234. If we increased liberalness by two units, the aforementioned interaction parameter of −17.241 would equal −17.241 + 1.007 + 1.007 = −15.227. As liberalness increases, the effect of the two-way interaction between ethnicity of the accused and ethnicity of the juror moves toward zero.

As in previous examples, it often helps readers to appreciate the three-way interaction by presenting relevant 2 × 2 tables at selected values of the continuous second-order moderator variable. This has been done in Table 3.6 for "low," "medium," and "high" values of liberalness-conservativeness, where a "low" value is defined as 1 *SD* below the mean liberalness-conservativeness value, a "medium" value is the mean liberalness-conservativeness value, and a "high" value is 1 *SD* above the mean liberalness-conservativeness value. We used the transformation strategy and the "recoding-of-dummy-variables" strategy to isolate all of the relevant statistics. The table is organized so that the focal independent variable is represented by rows and the first-order moderator variable by columns, and the second-order moderator is used to segregate the 2 × 2 tables. Beneath each 2 × 2 table is the interaction parameter. The three-way interaction is apparent from the fact that the two-way interaction parameter estimate varies across the selected levels of liberalness-conservativeness. The data can be presented graphically as a bar graph using the principles discussed earlier.

Qualitative Variables With More Than Two Levels

The above logic extends readily to the case of qualitative variables with more than two levels. One need keep in mind only the conditional nature of the coefficients and the particular cells of the design that the product term invokes. For example, in the first example, where the continuous variable was the focal independent variable, suppose we had three levels of ethnicity of the defendant (African American, European American, and Latino) and

TABLE 3.6
Cell Means and Two-Way Interaction Parameters for Guilt Attributions at Three Levels of Liberalness-Conservativeness

	Low L			*Medium L*			*High L*	
	AA Jur	*EA Jur*		*AA Jur*	*EA Jur*		*AA Jur*	*EA Jur*
AA def	49.18	80.92	AA def	40.23	56.24	AA def	31.27	31.56
EA def	50.07	48.31	EA def	40.23	39.00	EA def	30.39	29.71
	(49.18 – 50.07) – (80.92 – 48.31) = –17.24**			(40.23 – 40.23) – (56.24 – 39.00) = –0.97			(31.27–30.39) – (31.56 – 29.71) = –33.50**	

NOTE: AA = African American; EA = European American; def = defendant; Jur = juror; L = liberalness-conservativeness.
**$p < .05$.

three levels of the ethnicity of the juror (African American, European American, and Latino). The full equation would be

$$
\begin{aligned}
Y = a &+ b_1L + b_2AC_{\text{AA}} + b_3AC_{\text{L}} + b_4\,JU_{\text{AA}} + b_5\,JU_{\text{L}} \\
&+ b_6(L)(AC_{\text{AA}}) + b_7(L)(AC_{\text{L}}) + b_8(L)(JU_{\text{AA}}) \\
&+ b_9(L)(JU_{\text{L}}) + b_{10}(AC_{\text{AA}})(JU_{\text{AA}}) + b_{11}(AC_{\text{AA}})^*(JU_{\text{L}}) \\
&+ b_{12}(AC_{\text{L}})(JU_{\text{AA}}) + b_{13}(AC_{\text{L}})(JU_{\text{L}}) \\
&+ b_{14}(L)(AC_{\text{AA}})(JU_{\text{AA}}) + b_{15}(L)(AC_{\text{AA}})(JU_{\text{L}}) \\
&+ b_{16}(L)(AC_{\text{L}})(JU_{\text{AA}}) + b_{17}(L)(AC_{\text{L}})\,JU_{\text{L}} + e
\end{aligned}
\qquad [3.2]
$$

where L is the centered liberalness-conservativeness measure, AC_{AA} is a dummy variable for the ethnicity of the accused in which observations involving an African American defendant receive a 1 and all others receive a 0, AC_{L} is a dummy variable for the ethnicity of the accused in which observations involving a Latino defendant receive a 1 and all others receive a 0, JU_{AA} is a dummy variable for the ethnicity of the juror in which all African Americans receive a 1 and everyone else receives a 0, and JU_{L} is a dummy variable for the ethnicity of the juror in which all Latinos receive a 1 and everyone else receives a 0. The European American defendant is the reference group for the ethnicity of the accused variable and the European American jurors are the reference group for ethnicity of the juror variable. Although this equation may appear intimidating, it is easily processed using the principles discussed in this chapter. The omnibus three-way interaction is tested by applying the hierarchical test of Equation 1.1 to the comparison of the model in Equation 3.2 with a model that drops the predictors associated with b_{14} through b_{17}. The coefficient b_1 is the effect of liberalness-conservativeness on attributions of guilt for the European American defendant and European American jurors. Redefining the dummy variables and recalculating the equation to isolate different combinations of reference groups for the two moderator variables will yield the relevant statistics for the nine slopes defined by the 3×3 combination of ethnicity of the accused and ethnicity of the juror. In each case, the focus is on b_1.

The coefficient for any two-way term that includes L reflects a difference between slopes. For example, b_7 is the slope of guilt attributions on liberalness-conservativeness for the Latino defendant minus the same slope for the European American defendant when the focus is on only European American jurors. The coefficient for a given three-way term reflects a difference between slope differences. For example, b_{17} is the above slope difference for European American jurors subtracted from the same slope difference for Latino jurors. Although there are many terms with which to work, their interpretation is straightforward.

Summary

In sum, researchers often are interested in exploring interaction models that focus on three-way interactions. The predictor variables involved in the interaction might all be continuous or they might be a combination of continuous and qualitative variables. In either case, the researcher declares one of the variables the focal independent variable, a second variable is declared the first-order moderator variable, and a third variable is declared a second-order moderator variable. Appropriate product terms are calculated and then the interaction model is estimated including all of the "main-effect" terms as well as all of the product terms. The significance test of the omnibus interaction is reflected in the regression coefficient associated with the three-way product term when the omnibus interaction has a single degree of freedom. When the omnibus interaction has more than a single degree of freedom, the hierarchical test that compares a model with the three-way interaction terms with a model that excludes the three-way interaction terms is applied to evaluate the statistical significance of the overall interaction.

The strength of the interaction effect in unstandardized terms is indicated in the value of the regression coefficients reflecting the single-degree-of-freedom interaction contrasts. The strength of the effect in standardized terms is indexed by the squared semipart correlation for the three-way product term(s).

Interpretation of the interaction typically focuses on the regression coefficients associated with the focal independent variable and the regression coefficients associated with the various product terms that include the focal independent variable. Depending on which variables are the moderator variables, different centering and recoding strategies are used to help the reader appreciate the nature of the interaction.

4. ADDITIONAL CONSIDERATIONS

The present chapter considers a range of issues relevant to interaction analysis, most of which are somewhat more advanced than the issues discussed thus far.

Selected Issues

The Bilinear Nature of Interactions for Continuous Variables

As noted, when a continuous variable is part of an interaction, it is important to keep in mind that the traditional interaction model with product terms tests only for an interaction that has a specific form, namely, a bilinear interaction. Other forms of interaction may be operating, and exploratory analyses should be performed routinely to ensure that the correct type of interaction is being modeled. In the case of two continuous predictor variables, the classic product-term approach reflects a narrowly defined but probably widely applicable interaction form. As noted earlier, if X is the focal independent variable and Z is the moderator variable, the product-term approach models the coefficient for X as a linear function of Z. It is possible that the coefficient for X changes as a nonlinear function of Z, and if this is the case, the traditional product-term approach represents a misspecified model. A crude but sometimes informative way to explore this issue is to use a variant of bandwidth regression (Hamilton, 1992). In this approach, the moderator variable (Z) is grouped into 5 to 10 roughly equal-sized, ordered categories. The mean or median Z is calculated for each group. A regression analysis is then performed regressing Y onto X for each of the Z groups separately. Examination of the coefficients for Y on X across the 5 to 10 groups defined by Z should reveal a trend whereby the coefficient increases or decreases as a roughly linear function of the mean or median Z for each group across the groups. Stated another way, if one plots from such an analysis the Y on X coefficients against the mean (or median) Z values, a linear trend should be evident. If this is not the case, then a more complex interaction form may be needed.

Such complex interactions often can be modeled using product terms in the context of polynomial regression. For an introduction to polynomial analysis with interaction terms in multiple regression, see Jaccard, Turrisi, and Wan (1990). As one example, the steps for applying a model that assumes the coefficient for X is a quadratic function of Z rather than a linear function of Z, where both X and Z are continuous, are as follows:

1. Identify the focal independent variable, X, and the moderator variable, Z.
2. Make any desired transformations (e.g., mean center) on X and Z.
3. Calculate the square of the moderator variable, Z^2.
4. Calculate product terms between X and Z and between X and Z^2.
5. Fit the equation $Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 Z^2 + \beta_4 XZ + \beta_5 XZ^2 + \varepsilon$.

A hierarchical test for improvement in model fit by adding the XZ^2 term indicates whether the quadratic interaction effect is nontrivial (or one can simply examine the significance of the test associated with the parameter β_5). The coefficient for X at a given value of Z is defined by $\beta_1 + \beta_4 Z + \beta_5 Z^2$. The coefficient β_1 is the coefficient for X when $Z = 0$. One can transform Z (in Step 2 above) so that a score of 0 on the transformed variable takes on a theoretically meaningful value to isolate the relevant coefficient and confidence interval for the coefficient for Y on X at any given value of Z.

For the case involving qualitative and continuous predictor variables, assume Z is a dummy variable scored with 1s and 0s to represent group membership. In this case, Y is a nonlinear function of X for at least one of the groups, possibly both of them. Fit the following model: $Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 X^2 + \beta_4 XZ + \beta_5 X^2 Z + \varepsilon$. The effect of X on Y when $Z = 0$ is reflected by the quadratic model $\alpha + \beta_1 X + \beta_3 X^2$ within this equation. To find the effect of X on Y when $Z = 1$, recode Z by reverse coding it, recalculate the product terms, and rerun the computer program, again focusing on the resulting $\alpha + \beta_1 X + \beta_3 X^2$ terms in the equation.

Some methodologists (e.g., Ganzach, 1997) suggest that models that assume simple linear relationships and simple bilinear interactions are too restrictive and that the possibility of curvilinear effects should routinely be accommodated. One such approach is to fit a model that permits the emergence of either linear or quadratic relationships between Y and the predictors as well as interaction forms where the model will be sensitive to a coefficient for X that is either a linear or quadratic function of Z. For the case of continuous predictors, such a model would have the form

$$\begin{aligned} Y = \alpha &+ \beta_1 X + \beta_2 X^2 + \beta_3 Z + \beta_4 Z^2 + \beta_5 XZ \\ &+ \beta_6 XZ^2 + \beta_7 X^2 Z + \beta_8 X^2 Z^2 + \varepsilon \end{aligned} \qquad [4.1]$$

The effect of X at any given value of Z in this model can be isolated by applying the transformation strategy to Z so that the zero point corresponds to the value of interest and then focusing on the β_1 and β_2 coefficients that isolate the corresponding coefficients in the model $Y = \alpha + \beta_1 X + \beta_2 X^2$ when $Z = 0$. As discussed later, such a model has the advantage of helping to protect the theorist against detecting spurious interactions or missing true interactions. Critics would argue that the approach can overfit the data and forsakes the principle of parsimony in theory construction. Both arguments have merit. The model in Equation 4.1, of course, assumes that any nonlinearity is quadratic in form.

Calculating Coefficients of Focal Independent Variables at Different Moderator Values

In previous chapters, coefficients for the focal independent variable were calculated at different values of the moderator variable by either transforming the continuous moderator variable or redefining the reference group of a qualitative moderator variable and then rerunning the regression analysis on the computer. This approach, though cumbersome, has the advantage of producing the estimated standard errors and confidence intervals for all of the parameters of interest. Such confidence intervals are not readily calculated by hand (for relevant formulas, see Aiken & West, 1991). Occasions may arise where one wishes to calculate the coefficients from the initial equation without generating confidence intervals and without redoing the analyses with transformed variables. This section describes the general logic for doing so. We begin by showing the approach that was used to derive Equation 2.7. We then generalize the logic to other scenarios.

Consider the case where X is the focal independent variable and Z is the moderator variable in the equation

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon \qquad [4.2]$$

We want to determine the coefficient for X at some value of Z. We first isolate all terms on the right-hand side of the equation that contain X:

$$\beta_1 X + \beta_3 XZ$$

and then factor out the X as

$$X(\beta_1 + \beta_3 Z)$$

which yields the coefficient for X at any value of Z, namely,

$$\beta \text{ for } X \text{ at } Z = \beta_1 + \beta_3 Z \qquad [4.3]$$

For example, in Equation 4.2, if $\beta_1 = 1.2$ and $\beta_3 = 0.05$, then the coefficient for X when $Z = 2$ is $1.2 + (0.05)(2) = 1.30$. Note that when $Z = 0$, the value of the coefficient in Equation 4.3 is β_1, which underscores the point that β_1 is conditioned on Z being 0.

If X and Z are dummy variables, the logic of Equation 4.3 holds but is focused on only the relevant dummy variables. For example, suppose X has two dummy variables and Z has two dummy variables, yielding the following equation:

$$\begin{aligned} Y = \alpha &+ \beta_1 D_{X1} + \beta_2 D_{X2} + \beta_3 D_{Z1} + \beta_4 D_{Z2} + \beta_5 D_{X1} D_{Z1} \\ &+ \beta_6 D_{X1} D_{Z2} + \beta_7 D_{X2} D_{Z1} + \beta_8 D_{X2} D_{Z2} + \varepsilon \end{aligned}$$

Suppose we want to isolate the coefficient for the group scored 1 on D_{X1} versus the reference group on X for the case where $D_{Z1} = 1$ and $D_{Z2} = 1$. We first isolate only the terms and coefficients that directly involve D_{X1}:

$$\beta_1 D_{X1} + \beta_5 D_{X1} D_{Z1} + \beta_6 D_{X1} D_{Z2}$$

and factor out D_{X1} to yield

$$D_{X1}(\beta_1 + \beta_5 D_{Z1} + \beta_6 D_{Z2})$$

so that

$$\beta \text{ for } X \text{ at } D_{Z1} \text{ and } D_{Z2} = \beta_1 + \beta_5 D_{Z1} + \beta_6 D_{Z2}$$

In the case where $\beta_1 = 0.2$, $\beta_5 = 0.3$, $\beta_6 = 0.4$, $D_{Z1} = 1$, and $D_{Z2} = 1$, the coefficient for D_{X1} is $[-0.2 + (0.3)(1) + (0.4)(1)] = 0.90$.

Equations for three-way interactions use the same logic. In the case of three continuous predictors X, Q, and Z, the traditional interaction equation is

$$Y = \alpha + \beta_1 X + \beta_2 Q + \beta_3 Z + \beta_4 XQ + \beta_5 XZ \\ + \beta_6 QZ + \beta_7 XQZ + \varepsilon$$

The coefficient for X at a given combination of scores on Q and Z is

$$\beta \text{ for } X \text{ at } Q \text{ and } Z = \beta_1 + \beta_2 Q + \beta_5 Z + \beta_7 QZ$$

and the coefficient for XQ at a given value of Z is

$$\beta \text{ for } XQ \text{ at } Z = \beta_4 + \beta_7 Z$$

Partialing the Component Terms

It is sometimes stated that the product terms in regression equations represent interaction effects. By and of themselves, the product terms reflect an amalgamation of main effects and interactions. It is only when the component parts of the product term are included in the equation along with the product term that the orderly relationships described in this book emerge (coupled with an unconstrained intercept term). It is possible to model interactions in ways that lead one to exclude one or more of the component parts of the product term, but this typically represents interactions of a different form than those considered here.

Traditional interaction analysis uses what are called hierarchically well-formulated models. A hierarchically well-formulated (HWF) model is one in which all lower-order components of the highest-order interaction term are included in the model. For example, if interest is in a two-way interaction between X and Z, then an HWF model includes X, Z, and XZ as predictors. If interest is in a three-way interaction between Q, X, and Z, then an HWF model includes Q, X, Z, QX, QZ, XZ, and QXZ as predictors. For a qualitative predictor with dummy variables D_1 and D_2 and a continuous predictor Z, an HWF interaction model includes D_1, D_2, Z, $D_1 \times Z$, and $D_2 \times Z$. When an HWF model is used, the orderly relationships described in this book apply.

Of course, one can model certain types of interactions without using an HWF model. For example, a simple multiplicative model might have the form

$$Y = \alpha + \beta_1 XZ + \varepsilon$$

The fit (i.e., multiple correlation) of such a model is impacted by the metrics of X and Z. While in an HWF model a simple transformation that subtracts a constant from X does not affect model fit, this is not true in the above multiplicative model. If one's metric is arbitrary in the purely multiplicative model, then so is one's fit. Modeling interactions that do not involve HWF models can be a delicate enterprise.

Transformations

We have relied heavily on a simple transformation strategy (subtracting a constant from a measure) to isolate simple effects and their associated estimated standard errors, significance tests, and confidence intervals. These transformations can be used for the analysis of two-way interactions, three-way interactions, four-way interactions, or higher-order interactions. The analyst simply must use an HWF model and keep in mind the conditional nature of coefficients: Any time a variable, X, is involved in a product term, the coefficient associated with it alone is conditioned on the other variables in the product term being zero. Similarly, if a product term, XZ, is involved in a higher-order product term (e.g., XZQ), then the coefficient associated with XZ is conditioned on the other variables in the higher-order product term being zero ($Q = 0$).

When isolating a simple effect of X on Y at a given value of the moderator, Z, using the transformation strategy, it usually is best to work with the raw scores when applying the transformation to create the relevant zero point of interest on the moderator variable. Some researchers mean center the Z scores and then apply additional transformation strategies to these transformed Z scores (e.g., subtracting and adding a standard deviation to them).

Although this is legitimate, it has been our experience that researchers are sometimes surprised by what appears to be result reversal when using such doubly transformed scores. The algebra is tedious to demonstrate the underlying dynamics, so we only caution the reader that the safest practice is to avoid doubly transformed predictors and to always isolate theoretically desired values of the moderator by applying the transformation strategy to the original raw scores minus a constant that results in the zero point of interest.

Multiple-Interaction Effects

Consider a case where an investigator desires to model an outcome, Y, as a function of three continuous predictors, X, Q, and Z. The researcher does not expect a three-way interaction between the predictors but wants to evaluate all possible two-way interactions. There are multiple strategies that might be used. Some analysts perform a "chunk" test in which the fit of a model with all (two-way) interaction terms included is contrasted with the fit of a model with none of the interaction terms; that is, the interactions are tested as a "chunk" (Kleinbaum, 1992). The test is accomplished using Equation 1.1. If the difference in fit of the two models is trivial, then this suggests that none of the interaction terms are necessary and they are dropped from the model. If application of the "chunk" test reveals a nontrivial difference in model fit, then this suggests that at least one interaction term is important to retain. At this point, a hierarchical backward elimination strategy is used comparing the fit of a model that includes all of the interaction terms versus the fit of a model that drops a particular term of interest (vis-à-vis Equation 1.1). For example, if one is interested in evaluating the XZ interaction, one would compare the fit of the model

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 Q + \beta_4 XZ + \beta_5 XQ + \beta_6 QZ + \varepsilon$$

with the fit of the model

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 Q + \beta_4 XQ + \beta_5 QZ + \varepsilon$$

If the difference in fit between the models is trivial, then this suggests that the XZ term can be eliminated. However, if the difference in the fit of the model is nontrivial, then the term should be retained.

Some analysts systematically evaluate each interaction term in this fashion. Other analysts choose one term to focus on first, and if that term is eliminated, evaluate the remaining interaction terms with the previously eliminated term(s) expunged from the model. For example, if we tested XZ first for possible elimination and ultimately decided to drop it from the model, then the evaluation of QZ would focus on a backward elimination test where XZ was not present in the model; that is, we would evaluate

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 Q + \beta_4 XQ + \beta_5 QZ + \varepsilon$$

versus

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 Q + \beta_4 XQ + \varepsilon$$

The choice of which term to evaluate first for possible elimination is based sometimes on theoretical criteria, sometimes on whichever term has the largest p value associated with its regression coefficient in the full equation, or sometimes on both.

In multiple-interaction scenarios, there are many model-fitting criteria that can be invoked for the trimming of terms, and controversy exists about these strategies. In-depth consideration of the relevant issues is beyond the scope of this book. Interested readers are referred to Bishop, Feinberg, and Holland (1975), Hosmer and Lemeshow (1989), and Jaccard (1998) for a discussion of germane issues. The reader should be forewarned that seeming "anomalies" can occur as multiple-interaction terms of the same order are considered. For example, the "chunk" test might indicate that at least one of the product terms should be retained in the model, but the evaluation of each individual term may suggest that each term can be eliminated from the model. Or the "chunk" test may suggest that all of the terms be eliminated whereas evaluation of the individual terms may suggest otherwise. Or the results of the individual tests of one term may suggest that the term be retained and that all others be eliminated, but when the others are eliminated, the candidate for retention becomes nonsignificant and of marginal predictive value. How one deals with these scenarios depends on the theoretical questions being addressed, one's overarching statistical framework (e.g., null hypothesis testing, magnitude estimation, interval estimation), and the patterning of the data. In most analytic situations, the choice of terms to trim will be straightforward and noncontroversial, but this is not always the case.

When two separate interaction terms are included in the regression equation (e.g., for three continuous predictors, Q, X, Z, and both XZ and QZ are retained in the equation but no other interaction terms are), then the coefficient for a given interaction term is interpreted as described in previous chapters but with the proviso that the other two-way interaction (as well as all other covariates) is statistically held constant. The coefficient for any lower-order term is conditional on the other variables in *all* product terms with which it is involved being zero.

Standardized and Unstandardized Coefficients

The regularities for the regression coefficients discussed in this book apply to the unstandardized coefficients associated with the predictor variables. Although it is possible to use standardized coefficients in the analysis of interactions, such coefficients have the potential to lead theorists astray and will not exhibit the regularities we have noted. We generally recommend against their use, although occasions may arise that justify their analysis. As one illustration of the limitations of standardized coefficients, consider a simple bivariate regression where we regress a measure of income onto the number of years of education in order to determine the "value" of a year of education. The analysis is conducted in two different ethnic groups, African Americans and European Americans. Suppose that the analysis yielded identical standardized regression coefficients in the two groups, indicating that for every 1 *SD* that education changes, income is predicted to change 0.50 *SD*. One might conclude from this that the "value" of education is the same in the two groups. Suppose that the standard deviation for education is 3.0 in both groups but that for income it is 15,000 for European Americans and 6,000 for African Americans. Such a state of affairs yields unstandardized coefficients of 2,500 for European Americans and 1,000 for African Americans. Whereas for European Americans an additional year of education is predicted to be worth \$2,500, for African Americans, it is worth only \$1,000. There is a clear disparity between the groups that is not reflected in the standardized analysis.

The problem with the standardized analysis is that it creates different metrics for the two groups. Because the standard deviations are different, the metric of a standard deviation is different. The metric is in units of \$15,000 for the European Americans, but it is in units of \$6,000 for African Americans. Comparing groups on these different metrics is somewhat like measuring income in units of dollars for one group but units of British pounds for another group and then comparing groups without acknowledging the difference between the dollar and the pound. For a discussion of other limitations of standardized coefficients, see Jaccard, Turrisi, and Wan (1990).

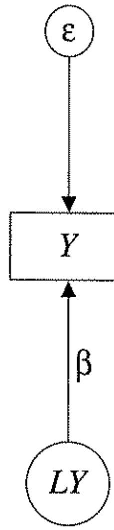

Figure 4.1 Path Diagram of a Measurement Model

Metric Properties

Metric properties are important for interaction analysis. The underlying issues are best understood by conceptualizing measurement in terms of classic latent variable modeling. An observed measure of a construct is viewed as an indicator of a latent variable that represents the true construct in question. In practice, we do not know a person's true score on this construct but instead use the observed measure to estimate it. The observed measure is influenced not only by the person's true standing on the latent variable but also by measurement error. A path diagram of the model is presented in Figure 4.1. If we assume a linear relationship between the observed measure and the latent variable, Figure 4.1 implies a regression equation in which the observed measure is regressed onto the latent variable:

$$Y = \alpha + \beta LY + \varepsilon$$

The above is a measurement model that describes how an observed measure maps onto a latent variable. Groups may differ on the intercept, slope, or error variance of this measurement model, and such differences can affect inferences about the true state of affairs regarding group differences on the latent variables (which are of primary theoretical interest). For example, two groups may have identical mean scores on the latent variable, but if they differ on their intercepts for the measurement model, the mean scores of the observed measures will be different. If two groups differ in the regression coefficient of the measurement model, then they can exhibit unequal variances on the observed measures even when they have equal variances on the latent variable. If two groups differ in their error variances, then the measures are differentially reliable for the groups (everything else being equal). The cleanest scenario analytically is where the different groups defined by the moderator variable have identical intercepts and slopes in the measurement model and the error variances are equal to zero or near zero. Deviations from this may introduce problems for interaction analysis. For a discussion of metric implications, methods for testing metric equivalence, and methods for metric adjustments, see Vandenberg and Lance (2000) and Busemeyer and Jones (1983).

Some methodologists assert that interaction analysis of the type discussed here is appropriate for only ratio-level measures. This is not the case. The methods can be used effectively for interval-level measures or measures that approximate reasonably well interval-level characteristics. This latter statement requires elaboration.

Some researchers erroneously refer to scales as being interval or ordinal in character. It is important to recognize that metric qualities are not inherent in scales but rather are inherent in data and hence are influenced by all of the facets of data collection. The extent to which a set of measures has interval properties is not dependent only on the scale used to make observations but also on the particular set of individuals on which the observations are made, the time at which the data are collected, the setting in which the data are collected, and so on. Consider the following simplistic yet pedagogically useful example. The height of five individuals is measured on two different metrics, inches and a rank order of height:

Individual	*Height (in inches)*	*Rank Order Height*
A	72″	5
B	71″	4
C	70″	3
D	69″	2
E	67″	1

As is well-known, the measures taken in inches have interval-level properties. For example, a difference of 1 between any two scores corresponds to the same physical difference on the underlying dimension of height. The actual height difference between individuals A and B corresponds to the same true underlying height difference between individuals C and D, and the metric reflects this (i.e., 72 – 71 = 1 and 70 – 69 = 1). Similarly, the difference between D and E is 69 – 67 = 2, and the difference between A and C is 2. These differences also reflect the same amount on the underlying dimension of height. Note, however, that these properties do not hold for the rank-order measure. The difference in scores between individuals A and B is 1 (i.e., 5 – 4) and the difference in scores for individuals D and E is also 1 (i.e., 2 – 1). These identical differences correspond to differing degrees of height disparities on the underlying dimension of height (i.e., the true difference between individuals D and E is larger than the true difference between individuals A and B, as is evident for the measure using inches). For these individuals, the rank-order measures have ordinal properties but not interval properties.

Now consider five different individuals with the following scores:

Individual	*Height (in inches)*	*Rank Order Height*
A	72″	5
B	71″	4
C	70″	3
D	69″	2
E	68″	1

Note that for these five individuals, the rank-order measures have interval-level properties. The difference in scores between individuals A and B is 1, as is the difference between individuals D and E. These differences correspond to the exact same amount on the underlying physical dimension. In this case, what we think of as traditionally being an ordinal "scale" actually yields measures with interval-level properties. Suppose that individual E was not 68 inches tall but instead was 67.9 inches tall. In this case, the rank-order measures are not strictly interval. But they are close and probably can be treated as if they are interval level without adverse effects.

This example illustrates that the crucial issue is not whether a set of measures is interval or ordinal. Rather, the issue is the extent to which a set of measures *approximates* interval-level characteristics. If the approximation is close, then the data often can be effectively analyzed using statistical methods that assume interval-level properties. If the approximation is poor, an alternative analytic strategy is called for. In this sense, interaction analyses can be performed on ordinal-level data as long as they approximate interval-level characteristics reasonably well.

Measurement Error

A topic that has received considerable attention in the statistical literature on interaction effects is the biasing effects of measurement error. It is well-known that unreliable measures can yield biased estimates of regression coefficients in multiple regression (e.g., Bohrnstedt & Carter, 1971). Social scientists frequently conduct research using fallible measures. Measurement error is thus a potential problem for the analysis of interaction effects involving continuous variables.

Using classical test theory, Busemeyer and Jones (1983) show that measurement error has the effect of attenuating hierarchical evaluations of product terms. The degree of attenuation is a direct function of the reliability of the product term, which we will call Pr. Under standard statistical assumptions, the amount of attenuation in changes in R^2 from the main-effect model will equal $\Pr(R_2^2 - R_1^2)$. For example, if the true incremental explained variance $(R_2^2 - R_1^2)$ due to the addition of a product term is .20 and the reliability of the product term is .70, then the observed incremental explained variance will equal (.20)(.70) = .14, everything else being equal.

Under certain reasonable statistical constraints,[2] if the true correlation between X and Z is 0, then the reliability of the product term XZ will equal the reliability of X times the reliability of Z. Thus, if one measure is relatively reliable ($r_{XX} = .80$) and the second measure is relatively unreliable ($r_{ZZ} = .50$), then the reliability of the product term will be lower than the reliability of the least reliable measure [(.80)(.50) = .40]. As the true correlation between X and Z increases, the reliability of the product term will increase, but not by much given the range of correlations and reliabilities typically observed in social science research (see Busemeyer & Jones, 1983, for elaboration). These facts underscore the difficulties that measurement error can create for tests of interaction, especially in situations with low statistical power. Using large sample sizes can offset the loss of power induced by measurement error for purposes of hypothesis testing, but a large N will *not* necessarily eliminate complications due to bias in the regression coefficients (e.g., Busemeyer & Jones, 1983; Evans, 1985).

Several resolutions to the problem of measurement error have been proposed. Cohen and Cohen (1975), Bohrnstedt and Marwell (1978), Heise (1986), and Fuller and Hidiroglu (1978) suggest approaches that require *a priori* knowledge of the reliabilities of the constituent variables. Cohen and Cohen (1983, p. 410) find fault with the method of correction that they suggested in their 1977 book in that the approach tends to overestimate the magnitude of the regression coefficients. The Bohrnstedt and Marwell approach has several limitations, detailed by Busemeyer and Jones (1983). Heise (1986) finds that his approach performed satisfactorily only under conditions where reliability was relatively high to begin with (e.g., above .90). The Fuller and Hidiroglu approach is promising but has been developed only for models without product terms. Extensions of this approach to product-term analysis would be valuable.

A second set of approaches to the problem of measurement error uses latent variable structural equation models (see Jaccard & Wan, 1996, for an introduction to these approaches). These strategies rely on multiple indicators of each construct to incorporate error theories into model tests and parameter estimation. Interaction models using these approaches are receiving a great deal of attention. A problem with many of the approaches is the need for large sample sizes and the assumption that the predictor variables all are normally distributed. Several strategies are being developed that relax the normality assumption. These include a two-stage least squares approach by Bollen (1996) and Bollen and Paxton (1998), quasi-maximum-likelihood-estimation approaches (Klein & Muthen, unpublished manuscript), and errors-in-variables factor-score approaches (Wall & Amemiya, 2000). At present, it is not possible to identify one analytic strategy that is necessarily superior to another across analytic scenarios. Although there is much work to be done, these methods hold promise.

The analyst who relies on traditional multiple regression for interaction analysis must acknowledge the potential bias due to measurement error, should use as valid and as reliable measures as possible, and should draw conclusions with appropriate caution. Ignoring measurement error is tantamount to assuming perfect reliability. This means that social scientists should devote considerable time and effort to developing high-quality measures before embarking on complex theory tests. The literature in psychometrics, questionnaire construction, and psychophysics is replete with well-established recommendations for reducing measurement error. For useful discussions of these practices, see Anderson (1981) and Wegenar (1982).

Robust Analyses and Assumption Violations

Thus far, we have presumed that the standard assumptions of inferential tests in ordinary least squares (OLS) regression hold true. We also have assumed that there are no outliers that mask the fundamental trends in the data. Unfortunately, the assumptions of OLS regression often are violated and inferential tests can be undermined accordingly. One approach to possible assumption violation is to use preliminary tests to evaluate the viability of the assumptions (e.g., a test of normality or a test of variance heterogeneity) and to alter the analytic strategy if the assumptions are violated. Many of these tests lack statistical power so that nonnormality or variance heterogeneity that matters remains undetected unless large sample sizes are used. In addition, the traditional *t* test and *F* tests of OLS regression were not derived under the scenario where a "screening test" is applied to determine whether to proceed with the analysis. Applying such screens can alter the shape of the underlying sampling distribution and cause more harm than good.

Another strategy is to transform the data in such a way that the data conform to the statistical model being fit to the data. Sometimes a transformation to alleviate one problem (nonnormality) creates other problems for the analysis (variance heterogeneity). Some transformations change the fundamental unit of the variables to the point that the new metric has no real-world meaning. This makes it difficult to work with the measure from a practical standpoint. Finally, the general philosophy of transformation puts the horse before the cart. Instead of manipulating data via transformations to conform to an underlying statistical model, why not use a statistical model that is appropriate for the data at hand?

Twenty years ago, finding such a statistical model might have proven to be an impossible challenge for some situations. However, with the advent of fast computers, far-reaching advances are being made in the field of robust statistical methods. Viable analytic methods are now available that do not make the strong assumptions of traditional statistical methods, that have excellent comparative statistical power, and that are outlier resistant. Some of these methods use the same principles described in this book (e.g., product terms that yield the same types of interpretations of coefficients) but do so in the context of outlier-resistant criteria that yield robust estimated standard errors and robust confidence intervals. Other methods involve entirely different frameworks, such as those based on smoothing. For useful expositions of robust methods, see Wilcox (1997, 2001). Wilcox (1997) explicitly includes a discussion of robust analyses for some forms of interaction.

Recent developments are merging robust methods with structural equation-modeling methods so that assumption violations, outliers, and measurement error all can be accommodated. These approaches include bootstrap estimation of standard errors and confidence intervals and concerted attention to distribution-free estimators (Arbuckle & Wothke, 1999). These techniques hold much promise for the future.

Within-Subject and Repeated-Measure Designs

Analytic scenarios arise where researchers desire to test the difference between slopes for within-subject or repeated-measure designs. One type of design is where an outcome variable is measured at two points in time and then regressed onto a common, stable predictor at each time period. For example, relationship satisfaction with one's parents might be regressed onto gender when respondents are in the seventh grade and again when they are in the ninth grade. Of interest is whether the effect of gender on relationship satisfaction in the seventh grade is different from the effect of gender on relationship satisfaction in the ninth grade. This design yields two equations:

$$Y_{t1} = \alpha_{t1} + \beta_{t1}X + \varepsilon_{t1}$$
$$Y_{t2} = \alpha_{t2} + \beta_{t2}X + \varepsilon_{t2}$$

where X is the stable predictor variable, Y_{t1} is the outcome variable measured at Time 1, and Y_{t2} is the outcome variable measured at Time 2. Judd, Kenny, and McClelland (2002) show that if one assumes that the path model in Figure 4.2 holds, then the null hypothesis of equal slopes in the two equations can be tested by applying traditional OLS regression, predicting the difference between Y_{t1} and Y_{t2} from X as follows:

$$Y_{t1} - Y_{t2} = a + bX + e \qquad [4.4]$$

The slope, b, in Equation 4.4 will equal the difference between b_{t1} and b_{t2}, and the test of significance of b evaluates the null hypothesis that $\beta_{t1} = \beta_{t2}$. Judd et al. (2002) discuss extensions of the test to designs with more than two repeated measures.

A second within-subject design focuses on the case where both X and Y vary over time, yielding the following two equations:

$$Y_{t1} = \alpha_{t1} + \beta_{t1}X_{t1} + \varepsilon_{t1}$$
$$Y_{t2} = \alpha_{t2} + \beta_{t2}X_{t2} + \varepsilon_{t2}$$

Of interest is the test of the null hypothesis that $\beta_{t1} = \beta_{t2}$, a test of what has been called "sequential moderation." James and Tetrick (1984) describe a T^2 statistic that can be used to test for sequential moderation that is based on a least squares regression model and that assumes that the model in Figure 4.3 holds.

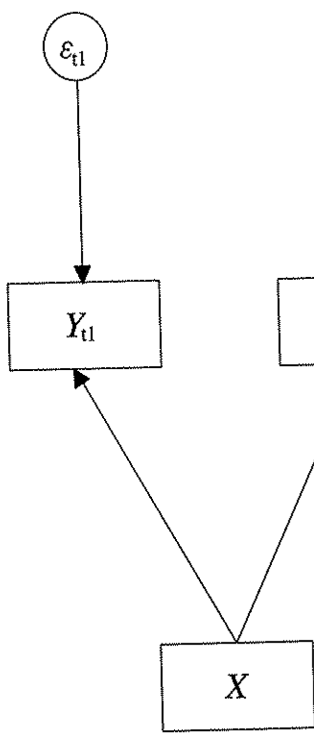

Figure 4.2 Model of Longitudinal Data With Stable Predictor

A limitation to these procedures is that the underlying models in Figures 4.2 and 4.3 may be incorrect, which can invalidate the tests. Often there will be correlated errors. Or the outcome at Time 1 may have a causal impact on the outcome at Time 2 independent of *X*. For the model in Figure 4.3, Y_{t2} may not be influenced only by X_{t2} but also by X_{t1} independent of X_{t2}. If the model is misspecified, then the tests of coefficient differences are undermined. Probably the best methods for testing interaction hypotheses of this type are those that use structural equation-modeling (SEM) frameworks (e.g., Kline, 1998). SEM can accommodate a wide range of causal models and error structures in the context of such tests, although they often require moderate to large sample sizes due to their reliance on asymptotic theory. Specialized methods in econometrics also may be of use (Greene, 1997).

Another common time-based interaction analysis is that based on growth curve models. In this approach, one specifies a mathematical function

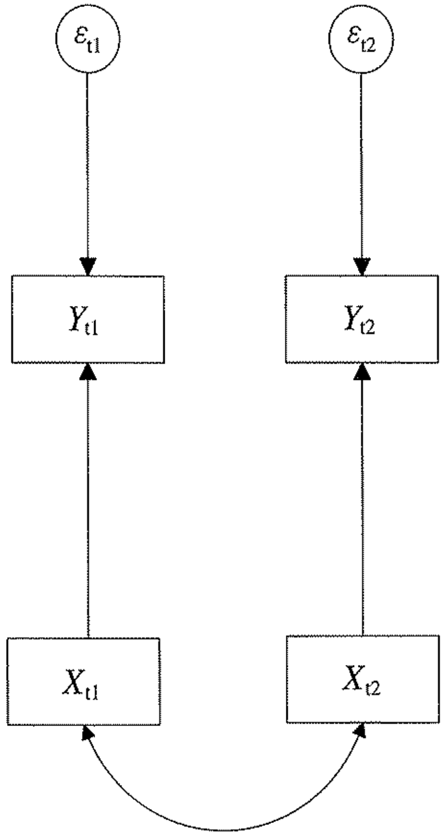

Figure 4.3 Model of Sequential Moderation

that describes how a variable changes over time for a given individual. For example, an individual's reading ability (Y) may change over the course of grades 1, 2, 3, 4, 5, and 6, and the change may be linear in form. The linear change can be described in terms of a slope and an intercept for that individual, with the slope indicating how much the reading level is predicted to change for the individual given a one-unit increase in time. If the units of time are years, then the slope indicates how much reading ability is predicted to change from one year to the next as one moves across grades 1 through 6. There likely will be individual differences in these slopes, with some children exhibiting steep slopes that indicate large changes in reading ability from year to year and other children exhibiting relatively flat slopes that indicate small changes in reading ability from year to year. Of interest are variables that predict the magnitude of these slopes. For example, do children who receive instruction in private schools, on average, exhibit steeper slopes than those who receive instruction in public schools? Does the magnitude of the slopes vary as a function of the child's socioeconomic standing? Such questions (which are inherently interaction based) are addressed using growth curve modeling as developed in the statistical literature on hierarchical linear modeling (Bryk & Raudenbush, 2002). Growth curve modeling also can be parameterized using a structural equation-modeling framework (Duncan, Duncan, Strycker, Li, & Alpert, 1999).

Ordinal and Disordinal Interactions

Social scientists distinguish between *ordinal* interactions and *disordinal* interactions. The distinction usually is made in the context of a qualitative predictor that has two or three groups coupled with a continuous predictor. A disordinal interaction is one in which the regression line that regresses Y onto the continuous predictor for one group intersects with the corresponding regression line for the other group. This also is referred to as a crossover interaction. An ordinal interaction is one in which the regression lines are nonparallel but they do not intersect. Figure 4.4 presents an example of a disordinal and an ordinal interaction.

Statisticians have expressed some wariness about ordinal interactions. Such interactions, they contend, may be an artifact of the metric of the dependent variable. Nonparallel regression lines frequently can be made parallel by means of a monotonic transformation of the Y scores. If the metric intervals of Y are truly arbitrary, then it makes sense from the standpoint of scientific parsimony to perform such transformations and remove the false moderator effect. Ordinal interactions, however, should not be dismissed if their metrics are meaningful. As Cronbach and Snow (1981) demonstrate, such interactions can be substantively important and, when coupled with cost-benefit criteria, can be crucial for classification decisions.

For any given pair of nonparallel regression lines, there is always a point where the lines will intersect. In this sense, all interactions are disordinal in theory. Interactions are classified as being ordinal if, *within the range of scores being studied* (e.g., for IQ scores between 90 and 110), the regression lines do not intersect. Consider the case of two groups, each of which can be described by its linear equation of Y on X. It is possible to identify the point on the continuous predictor where the regression lines for the two groups intersect using the following formula:

$$P_{\mathrm{I}} = (a_1 - a_2)/(b_2 - b_1)$$

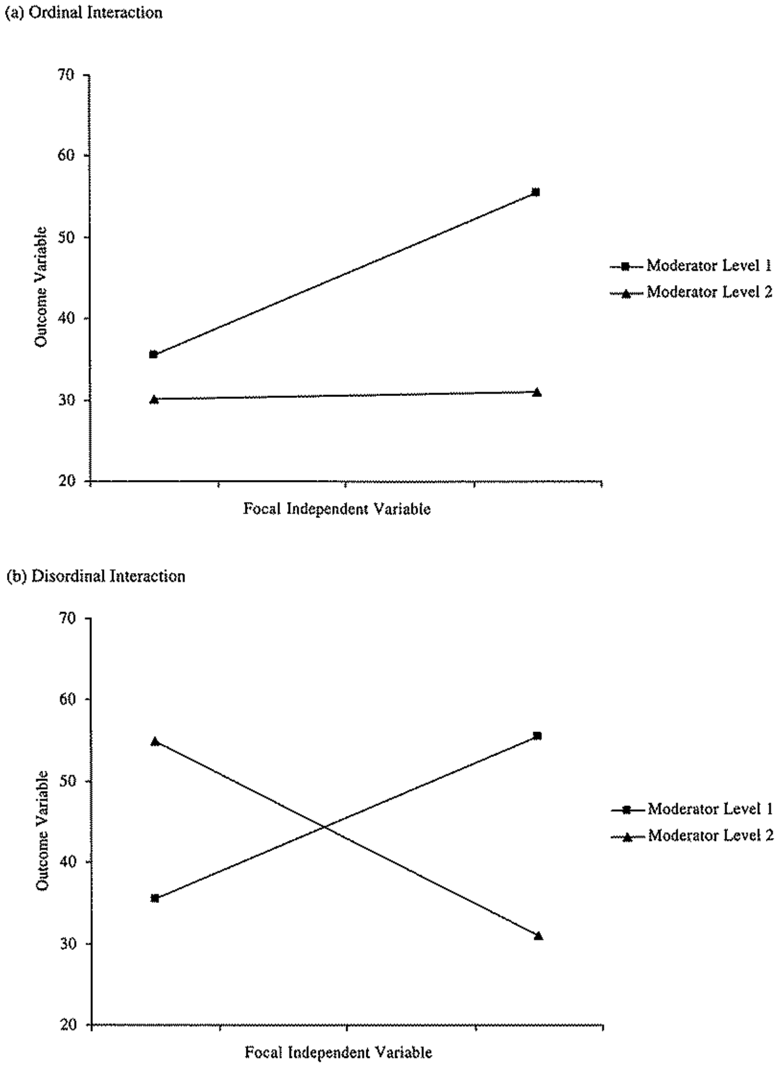

Figure 4.4 Ordinal and Disordinal Interactions

where a_1 is the intercept for the first group based on the regression of the outcome variable on the continuous predictor for that group, a_2 is the intercept for the second group based on the regression of the outcome variable on the continuous predictor for that group, b_1 is the slope for the first group based on the regression of the outcome variable on the continuous predictor for that group, and b_2 is the slope for the second group based on the regression of the outcome variable on the continuous predictor for that group.

As an illustration of the relevant calculations and a substantive application, consider the following example: A psychologist is interested in the relative effects of two types of therapy on childhood self-esteem. An experimental study is conducted where children receive an intervention designed to raise their self-esteem. Half of the children receive intervention A and the other half receive intervention B. Assignment to groups is random. No control group is included in the study because previous research indicates that both interventions are effective in raising childhood self-esteem relative to a no-treatment control. How children respond to the two treatments is thought to be moderated by the quality of their relationship with their parents. For children who have relatively positive relationships with their parents, it is thought that intervention A will be more effective than intervention B. For children who have relatively poor relationships with their parents, it is thought that intervention B will be more effective than intervention A. The type of intervention received was dummy coded using a single dummy variable, and the quality of the relationship between the child and the parent was measured on a 51-point scale (ranging from 0 to 50) based on a clinician's rating. The clinician had spent considerable time with both the parent and the child and made a rating based on these interviews. Higher scores indicated a better-quality relationship. Self-esteem was measured on a scale from 0 to 20, with higher scores indicating higher levels of self-esteem. In this study, postintervention self-esteem is the outcome variable, the type of intervention is the focal independent variable, and the quality of the parent-child relationship is the moderator variable. The self-esteem scores were regressed onto the dummy variable, the quality of the relationship measure, and the product of the two variables (without any mean centering in the analysis). The interaction effect was statistically significant.

We begin by calculating the separate regression equations for the two groups that regress the self-esteem scores onto the quality of relationship scores. These can be isolated using the procedures discussed in Chapter 2 and are

$$\text{Intervention A: } Y = 1.193 + 0.098Z + e$$
$$\text{Intervention B: } Y = 7.193 + -0.107Z + e$$

A plot of the scores and the regression lines would reveal that the interaction is disordinal. The point of intersection is

$$P_{\mathrm{I}} = (1.193 - 7.193)/(-0.107 - 0.098) = 29.27$$

This point describes the score on Z where the predicted self-esteem scores are the same for the two groups. Thus, when the quality of the relationship corresponds to a score of 29.27, the postintervention self-esteem scores of the children are predicted to be the same in the two intervention conditions. As the quality of the relationship exceeds 29.27, the postintervention self-esteem scores are predicted to be higher in intervention A than in intervention B. As the quality of the relationship falls below 29.27, the postintervention self-esteem scores are predicted to be higher in intervention B than in intervention A.

The results of this analysis suggest a basis for placement decisions for the children. If on a preintervention measure their quality of relationship with their parent is less than 29.27, then they should be given intervention B. If on a preintervention measure their quality of relationship with their parent is greater than 29.27, then they should be given intervention A.

Cronbach and Gleser (1957) review the logic of such treatment and classification decisions in educational, organizational, and psychological research. These authors note that decisions about the assignment of people to treatments (e.g., clinical interventions, type of educational curricula, type of job) are frequently guided by the identification of crossover points in disordinal interactions: Persons to the right of the crossover point are assigned to one treatment, and persons to the left of the crossover point are assigned to the other treatment. In contrast, ordinal interactions imply the same treatment might be used for all individuals.

Regions of Significance

Although the above analysis of point of intersection is useful, we must recognize that there is sampling error that affects our specification of a point of intersection. Potthoff (1964) developed a method, based on the classic work of Johnson and Neyman (1936), that establishes "regions of significance" relevant to the interpretation of an intersection point in a disordinal interaction. The focus of this approach is where the researcher has a qualitative focal independent variable and a continuous moderator variable. For the case of two groups, the technique defines a *range* of scores on the moderator variable where the members of one group on the focal independent variable are not expected to have a higher Y score than the members of the other group taking into account sampling error (see Aiken & West, 1991, for a more formal characterization of these intervals). In the example from the previous section, applying the procedures to be described below yielded a region with values ranging from 27.36 to 31.07. This means that when the quality of the parent-child relationship exceeds 31.07, we can be reasonably confident (based on a 95% confidence level) that the predicted postintervention self-esteem for the group receiving intervention A exceeds that of the group receiving intervention B. When the quality of the parent-child relationship is less than 27.36, we can be reasonably confident that the predicted postintervention self-esteem of the group receiving intervention B exceeds that of the group receiving intervention A.

The relevant values are obtained by solving for the two values (which we call *CV*) in the expression

$$CV = [-B \pm (B^2 - AC)^{1/2}]/A \qquad [4.5]$$

To define the values of *A*, *B*, and *C*, let Y = the outcome variable, Z = the continuous moderator variable, N = the total number of individuals across both groups that define the focal independent variable, n_1 = the number of individuals in group 1, n_2 = the number of individuals in group 2, F_α = a tabled *F* ratio with 2 and $N - 4$ *df* that corresponds to the experimenter-defined critical *F* value associated with an *a priori* specified alpha level (traditionally .05), M_1 = the mean score on *Z* for group 1, M_2 = the mean score on *Z* for group 2, S_1 = the sum of squares regression for *Z* for group 1, S_2 = the sum of squares regression for *Z* for group 2, E = the residual sum of squares for the regression of *Y* on *Z* for group 1 plus the corresponding residual sum of squares for the regression of *Y* on *Z* for group 2, a_1 = the intercept for the linear regression of *Y* onto *Z* for group 1, a_2 = the intercept for the linear regression of *Y* onto *X* for group 2, b_1 = the slope for the linear regression of *Y* onto *Z* for group 1, b_2 = the slope for the linear regression of *Y* onto *Z* for group 2. Then

$$A = [-2F_\alpha/(N-4)]E[1/S_1 + 1/S_2] + (b_1 - b_2)^2$$
$$B = [2F_\alpha/(N-4)]E[M_1/S_1 + M_2/S_2] + (a_1 - a_2)(b_1 - b_2)$$
$$C = [-2F_\alpha/(N-4)]E[N/(n_1 n_2) + M_1^2/S_1 + M_2^2/S_2] + (a_1 - a_2)^2$$

The within-group regression equations should not use centered measures of *Z*. For greater discussion of this method, relevant computer code, and extensions to more complex scenarios, see Aiken and West (1991). Cronbach and Snow (1981) discuss the strengths and limitations of these approaches.

Confounded Interactions

Some analysts have noted that interaction effects often are confounded with curvilinear effects of X on Y, thereby complicating interaction analysis (e.g., Lubinski & Humphreys, 1990). Data may be the result of a generating process that derives from a curvilinear relationship between X and Y, but when an interaction model is fit to the data using X, Y, and a third variable, Z, a false interaction results.

To illustrate the basic ideas, consider the case where the true underlying model of the relationship between Y and X is curvilinear and quadratic in form as expressed by the following model:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \varepsilon$$

An interaction model, by contrast, is

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

It turns out that as the correlation between X and Z increases, the correlation between X^2 and XZ also increases. This means that when X and Z are correlated, there will be confounding between the two models (Busemeyer & Jones, 1983; Lubinski & Humphreys, 1990). Depending on the nature of the quadratic form, fitting an interaction model when the underlying relationship is curvilinear can result in spurious conclusions about interactions. If both interactive and curvilinear effects are operative, then fitting the traditional interaction model can result in missed interactions, spurious interactions, or misleading interactions that are opposite in sign to the true interaction effect (Ganzach, 1997).

The essence of the problem is that of fitting misspecified models to the data. Doing so can lead the theorist astray. Researchers need to think carefully about the possible models that can account for data and then explore these models accordingly. If a curvilinear effect is not theoretically plausible and makes no conceptual sense, then it should not be pursued. If such a model is theoretically viable, then it should be considered. It may be the case that an interaction term becomes statistically nonsignificant when covariates mapping onto curvilinear effects are introduced into the equation. This does not necessarily invalidate the interaction model. It only suggests that an alternative model also can account for the data. The bottom line is that common sense and theory need to frame the types of models explored and that researchers must recognize that multiple models may need to be considered and contrasted before settling on an interaction model. The work of Lubinski and Humphreys (1990) as well as others suggests that researchers should consider the viability of curvilinear models as competing alternative models when contemplating interaction effects.

Optimal Experimental Designs and Statistical Power

Several analysts have lamented the difficulty of detecting interaction effects in linear models because of low statistical power. McClelland and Judd (1993) explore a host of reasons that underlie low statistical power in field settings. These researchers note that the statistical power to detect an interaction effect is heavily dependent on the nature of the distributions of the component variables of the product term (i.e., X and Z for the product term XZ). McClelland and Judd (1993) suggest design strategies based on the oversampling of extreme cases of X and Z that can be used when practical constraints dictate a small sample size and statistical power for detecting an interaction is expected to be low. These methods must be used with caution, however, because they can yield biased estimates of standardized effect sizes, such as estimates of the increment in squared multiple correlations based on the additive versus the interactive model.

When conducting statistical power analyses for interaction effects, it is important to take into account theoretical limits on the effect size of the interaction. For example, traditional power analysis for an interaction term is based on the incremental explained variance that results from adding a product term to a "main-effect-only" model. If theory dictates *a priori* an ordinal interaction between the two predictors and the interaction effect in the population is nonzero, then, by definition, the squared multiple correlation for the main-effect-only population model must be nonzero. The larger the effect size of the ordinal interaction, the larger will be the effect size for the main-effect-only model, which is used as the baseline from which to evaluate the statistical power of the interaction term. The formal relationship between these two effect sizes is described in Rogers (2002). Such linkages must be respected when conducting power analysis for ordinal interactions because researchers may unwittingly apply power analyses that specify a standardized interaction effect size in the population that is theoretically impossible. For more details, see Rogers (2002).

Covariates

Covariates can be added to any of the regression models discussed in this book without disrupting the interpretation of the coefficients involved in the interaction. For example, a covariate, Q, might be added to the product-term model as follows:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_4 Q + \varepsilon$$

In this model, β_3 is the number of units that the effect of X on Y is predicted to change given a one-unit change in Z, holding Q constant. The intercept is the predicted mean Y when X, Z, and Q all equal 0. Simple linear transformations of Q do not affect the coefficients of the other predictors, but they do affect the intercept term as the zero point of Q changes.

Control for Experimentwise Error

In some interaction analyses, multiple single-degree-of-freedom contrasts are pursued. When this is the case, the per-comparison alpha level remains at the specified alpha for a given contrast (usually .05), but the probability of at least one Type I error occurring across the set of contrasts exceeds the per-comparison alpha. In such instances, some researchers invoke statistical adjustments that maintain the experimentwise alpha level (i.e., the probability of obtaining at least one Type I error across a set of contrasts) at a specified level across the contrasts. The most popular method for doing so is the traditional Bonferroni procedure, although the technique is conservative. More powerful alternatives can be used that effectively control the Type I error rate (see Westfall et al., 1999). As one example, Holm (1979; see also Holland & Copenhaver, 1988; Seaman, Levin, & Serlin, 1991) has suggested a sequential modified Bonferroni method. Here is how it is applied. First, a p value is obtained for each contrast in the family of contrasts. The p values are then ordered from smallest to largest. If two p values are identical, they are ordered arbitrarily or using theoretical criteria. The contrast with the smallest p value is evaluated against an alpha of $.05/k$, where k is the total number of contrasts in the family. If this leads to rejection of the corresponding null hypothesis (because the observed p value is less than the adjusted α), then the next-smallest p value is tested against an alpha level of $.05/(k - 1)$, where $k - 1$ is the remaining number of contrasts. If this test leads to null hypothesis rejection, then the next-smallest p value is tested against an alpha level of $.05/(k - 2)$, and so on, until a nonsignificant difference is observed. Once a statistically nonsignificant difference is observed, all remaining contrasts are declared nonsignificant.

When the reported single-degree-of-freedom contrasts are merely illustrative and designed to provide the reader with a sense of the interaction, then the invocation of experimentwise controls probably is unnecessary. If one is going to make a theoretical statement tied to the analysis, then the issue is more germane. Decisions to invoke experimentwise error controls are complex and governed by a wide range of issues, including statistical power and the consequences of Type I and Type II errors.

Omnibus Tests and Interaction Effects

A common strategy used in interaction analysis is to first perform an omnibus test of an interaction effect and then to pursue single-degree-of-freedom interaction contrasts only if the omnibus effect is statistically significant. The omnibus test is used as a basis for protecting against inflated Type I errors across the single-degree-of-freedom interaction contrasts underlying the omnibus interaction. In general, most such two-step approaches have been discredited as an effective means of controlling experimentwise error rates (Jaccard, 1998; Wilkinson, 1999). An alternative strategy is to move directly to the single-degree-of-freedom contrasts that are of theoretical interest and to invoke controls for experimentwise error (e.g., the modified Bonferroni test) at that level independent of the results of an omnibus test. This does not mean that omnibus tests of interactions never will be meaningful. Such tests may be of interest if one wants to document the effect size of an overall interaction between two or more variables. In addition, if the omnibus interaction is not even remotely close to attaining statistical significance, then it is unlikely that any of the interaction contrasts will be significant. The omnibus test thus can be an effort-saving device. For further discussion of this topic, see Jaccard (1998).

Some Common Misapplications

Instances exist in the literature of poor practices with respect to interaction analysis. One we mentioned earlier. It involves the interpretation of regression equations computed in two or more groups separately and then declaration of group differences without formally testing those differences. Another practice is interaction analysis that reduces a continuous predictor variable to a two-valued indicator through the use of a median split. This strategy often is invoked so that interactions can be explored using traditional analysis of variance. Such practices are undesirable because they throw away useful information, they often result in less statistical power, and they can introduce false effects (Maxwell & Delaney, 1993). Interactions between continuous and qualitative predictors can be analyzed effectively using the general linear model without recourse to reducing the continuous predictor to a crude, bilevel indicator.

Interaction Models With Clustered Data and Random Coefficient Models

Some research designs involve clustered data where the researcher is interested in exploring the effects of cluster characteristics on the slope of Y on X. As an example, consider the case where data are collected on 5,000 students, 100 in each of 50 different schools. The researcher is interested in the effects of peer pressure (X) on drug use (Y) and whether this varies as a function of the size of the student body of the school (Z). One way of conducting this analysis is first to record a value of Z for a given individual the size of the school that the student attends (so that all students from the same school receive the same Z score). Then the XZ product term is formed and an OLS analysis regressing Y onto X, Z, and XZ based on $N = 5{,}000$ is performed. This strategy is problematic. One limitation is that the residuals in the population may not be independent. Students from the same school often are more alike one another as compared with students from other schools, and such dynamics may conspire to introduce dependencies in residuals. The clustering that is operative must be taken into account, and the regression strategy described does not do so.

In such scenarios, statisticians often apply a statistical model that is different from traditional OLS and that is called *random coefficient regression* or *hierarchical linear modeling* (HLM). Suppose we posit that the variability in regression coefficients across schools can be modeled by the following equation:

$$\beta_j = \alpha + \beta Z_j + \varepsilon_j \qquad [4.6]$$

where β_j is the regression coefficient for Y on X for school j, Z_j is the school size for school j, α is an intercept term for the regression of the values of the β_j onto the values of Z_j, β is the slope for the regression of the values of the β_j onto the values of Z_j, and ε_j is the error term from the regression of the values of the β_j onto the values of Z_j. In general, the ε_j are assumed to be normally distributed with a mean of zero and a constant variance at given values of Z. The presence of ε_j in Equation 4.6 is the defining feature of the random coefficient regression model. Such a model is distinct from traditional regression models with product terms, and special analytic methods are required to estimate α and β in Equation 4.6 and test the significance of the estimates. Random coefficient regression models inherently focus on interactions (because the effect of Y on X varies with Z), but they usually are applied in the context of clustered data. For a discussion of these methods, see Bryk and Raudenbush (2002).

Continuous Versus Discrete Predictor Variables

Some examples in this book used many-valued, quantitative, discrete variables as predictors and treated them as if they were continuous variables. This is not an unreasonable approach as long as such predictors have many values and behave in a way that roughly conforms to the assumptions of OLS regression. If a discrete variable has few values, then one might consider using dummy variables to represent it. Alternatively, one can use methods to model the effects of discrete predictor variables that are not regression based (Bollen, 1989; Joreskog & Sorbom, 1993).

A related issue is one where the underlying latent variable for a construct is continuous but the observed measure that is used by the researcher is discrete in character. This is often the case when rating scales are used (e.g., a rating scale that ranges from 1 to 7 is used to reflect the underlying continuous construct of self-esteem). In general, the analyst will not encounter problems in such cases as long as the measure has many values and behaves in a way that roughly conforms to the assumptions of OLS regression when used in traditional regression models. Analytic methods for overcoming problems resulting from too coarse a measure are discussed in Bollen (1989) and Joreskog and Sorbom (1993).

The Moderator Framework Revisited

The moderator framework to interaction analysis is useful in that it provides a tool by which interactions can be interpreted at a substantive level. Some statisticians object to the approach because it conceptually minimizes the symmetry of effects that operates in interaction models and because it focuses attention away from some of the coefficients in the interaction model. These objections have merit but often are offset by the conceptual gains of framing the interaction in a way that makes theoretical sense. Non-moderator-based conceptions of interactions emphasize the concept of nonadditivity and the isolation of effects based on residualized means (e.g., Rosnow & Rosenthal, 1996). Neither the moderator approach nor these latter approaches are necessarily superior to the other. They simply are different ways of viewing the data. Some methodologists (e.g., Pedhazur, 1997) define interactions not only in terms of the underlying statistical model but also in terms of explicit features of the research design coupled with the patterning of data in that design. These definitions are esoteric and unnecessarily limit interaction analysis. It is, of course, important to take into account design features when thinking about the meaning of product-term analyses. There are scenarios where such analyses can be construed as reflecting mediation rather than moderation. But for a wide range of applications, the framework outlined here will prove to be a useful way to conceptualize and approach interaction analysis.

REFERENCES

AIKEN, L., and WEST, S. (1991). *Multiple Regression*. Newbury Park, CA: Sage.

ANDERSON, N. H. (1981). *Methods of Information Integration Theory*. New York: Academic Press.

ARBUCKLE, J. L., and WOTHKE, W. (1999). *AMOS 4.0 User's Guide*. Chicago: Smallwaters.

BERRY, W. D., and FELDMAN, S. (1985). *Multiple Regression in Practice*. Beverly Hills, CA: Sage.

BISHOP, Y. M., FEINBERG, G., and HOLLAND, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

BOHRNSTEDT, G. W., and CARTER, T. M. (1971). "Robustness in regression analysis," pp. 118–146 in H. L. Costner (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

BOHRNSTEDT, G. W., and MARWELL, G. (1978). "The reliability of products of two random variables," pp. 254–273 in K. F. Schuessler (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

BOLLEN, K. A. (1996). "An alternative two stage least squares (2SLS) estimator for latent variable equations." *Psychometrika*, 61: 109–121.

BOLLEN, K. A., and PAXTON, P. (1998). "Interactions of latent variables in structural equation models." *Structural Equation Modeling*, 5: 267–293.

BRYK, A., and RAUDENBUSH, S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

BUSEMEYER, J. R., and JONES, L. (1983). "Analysis of multiplicative combination rules when the causal variables are measured with error." *Psychological Bulletin*, 93: 549–562.

COHEN, J., and COHEN, P. (1975). *Applied Multiple Regression for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

COHEN, J., and COHEN, P. (1983). *Applied Multiple Regression for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.

CRONBACH, L. (1987). "Statistical tests for moderator variables: Flaws in analysis recently proposed." *Psychological Bulletin*, 102: 414–417.

CRONBACH, L. J., and GLESER, G. C. (1957). *Psychological Tests and Personal Decisions (2nd ed.)*. Urbana: University of Illinois Press.

CRONBACH, L. J., and SNOW, R. E. (1981). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington.

DUNCAN, T. E., DUNCAN, S. E., STRYCKER, L. A., LI, F., and ALPERT, A. (1999). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications*. Mahwah, NJ: Erlbaum.

EVANS, M. T. (1985). "A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis." *Organizational Behavior and Human Decision Processes*, 36: 305–323.

FRIEDRICH, R. (1982). "In defense of multiplicative terms of multiple regression equations." *American Journal of Political Science*, 26: 797–833.

FULLER, W. A., and HIDIROGLU, M. A. (1978). "Regression estimation after correction for attenuation." *Journal of the American Statistical Association*, 73: 99–104.

GANZACH, Y. (1997). "Misleading interaction and curvilinear terms." *Psychological Methods*, 2: 235–247.

GREENE, W. H. (1997). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

HAMILTON, L. C. (1992). *Regression With Graphics: A Second Course in Applied Statistics*. Belmont, CA: Brooks-Cole.

HARDY, K. (1993). *Regression with Dummy Variables*. Newbury Park, CA: Sage.

HEISE, D. R. (1986). "Estimating non-linear models: Correcting for measurement error." *Sociological Methods of Research*, 14: 447–472.

HOLLAND, B. S., and COPENHAVER, M. (1988). "Improved Bonferroni-type multiple testing procedures." *Psychological Bulletin*, 104: 145–149.

HOLM, S. (1979). "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics*, 6: 65–70.

HOSMER, D. W., and LEMESHOW, S. (1989). *Applied Logistic Regression*. New York: Wiley.

JACCARD, J. (1998). *Interaction Effects in Factorial Analysis of Variance*. Thousand Oaks, CA: Sage.

JACCARD, J., and DODGE, T. (2003). "Specification of contingent effects in linear models," in M. Hardy and A. Bryman (eds.), *Handbook of Data Analysis*. Thousand Oaks, CA: Sage, in press.

JACCARD, J., TURRISI, R., and WAN, C. (1990). *Interaction Effects in Multiple Regression*. Newbury Park, CA: Sage.

JACCARD, J., and WAN, C. (1996). LISREL *Approaches to Interaction Effects in Multiple Regression*. Thousand Oaks, CA: Sage.

JAMES, L., and TETRICK, L. (1984). "A multivariate test of homogeneity of regression weights for correlated data." *Educational and Psychological Measurement*, 44: 769–780.

JOHNSON, P. O., and NEYMAN, J. (1936). "Tests of certain linear hypotheses and their application to some educational problems." *Statistical Research Memoirs*, 1: 57–93.

JORESKOG, K., and SORBOM, D. (1993). *LISREL VIII*. Chicago: Scientific Software.

JUDD, C. M., KENNY, D. A., and McCLELLAND, G. (2002). "Estimating and testing mediation and moderation in within subject designs." *Psychological Methods*, 6: 115–134.

KLEINBAUM, D. G. (1992). *Logistic Regression: A Self Learning Text*. New York: Springer.

KLINE, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press.

LEWIS-BECK, M. S. (1980). *Applied Regression: An Introduction*. Beverly Hills, CA: Sage.

LUBINSKI, D., and HUMPHREYS, L. (1990). "Assessing spurious 'moderator effects': Illustrated substantively with the hypothesized ('synergistic') relation between spatial and mathematical ability." *Psychological Bulletin*, 107: 385–393.

MAXWELL, S., and DELANEY, H. (1993). "Bivariate median splits and spurious statistical significance." *Psychological Bulletin*, 113: 181–190.

McCLELLAND, G. H., and JUDD, C. M. (1993). "Statistical difficulties of detecting interactions and moderator effects." *Psychological Bulletin*, 114: 376–390.

PEDHAZUR, E. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction*. New York: Harcourt Brace.

POTTHOFF, R. F. (1964). "On the Johnson-Neyman technique and some extensions thereof." *Psychometrika*, 29: 241–256.

ROGERS, W. M. (2002). "Theoretical and mathematical constraints of interactive regression models." *Organizational Research Methods*, 5: 212–230.

ROSNOW, R. L., and ROSENTHAL, R. (1996). "Contrasts and interaction effects redux: Five easy pieces." *Psychological Science*, 7: 253–257.

SCHROEDER, L. D., SJOQUIST, D. L., and STEPHAN, P. (1986). *Understanding Regression Analysis: An Introductory Guide*. Beverly Hills, CA: Sage.

SEAMAN, M. A., LEVIN, K. R., and SERLIN, R. C. (1991). "New developments in pairwise multiple comparisons: Some powerful and practicable procedures." *Psychological Bulletin*, 110: 577–586.

VANDENBERG, R. J., and LANCE, C. (2000). "A review and synthesis of the measurement in variance literature: Suggestions, practices, and recommendations for organizational research." *Organizational Research Methods*, 3: 4–69.

WALL, M., and AMEMIYA, Y. (2000). "Estimation for polynomial structural equation models." *Journal of the American Statistical Association*, 95: 929–940.

WEGENAR, B. (1982). *Social Attitudes and Psychological Measurement*. Hillsdale, NJ: Lawrence Erlbaum.

WESTFALL, P. H., TOBIAS, R., ROM, D., WOLFINGER, R., and HOCHBERG, R. D. (1999). *Multiple Comparisons and Multiple Tests Using the SAS System*. Cary, NC: SAS Institute.

WILCOX, R. R. (1997). *Introduction to Robust Estimation and Hypothesis Testing*. New York: Academic Press.

WILCOX, R. R. (2001). *Fundamentals of Modern Statistical Methods*. New York: Springer-Verlag.

WILKINSON, L. (1999). "Statistical methods in psychology journals: Guidelines and explanations." *American Psychologist*, 54: 594–604.

NOTES

1. When suppressor variables are present, this interpretation of semipart correlations is flawed (see Cohen & Cohen, 1983).

2. The assumptions are (a) that the residuals in the population are normally distributed, (b) that the latent continuous variables are multivariately normally distributed (but not the latent product term), and (c) that a given observed score is a function of a true score and an error score in accord with classic test theory.

ABOUT THE AUTHORS

James Jaccard is Distinguished Professor of Psychology at the University of Albany, State University of New York. His primary research interests are in the areas of the psychology of population dynamics and adolescent risk behavior, with an emphasis on adolescent unintended pregnancy and adolescent drunk driving. His work has focused on family-based approaches to dealing with adolescent problem behaviors. He has authored three other monographs on interaction effects in the Sage series.

Robert Turrisi is Professor of Psychology at Boise State University and the Director of the Family Studies Research program. His research focuses on the application of psychological theories of behavioral decision making to adolescent health issues and family relationships. This includes developing and evaluating the efficacy of large-scale parent-based interventions to prevent college student binge drinking, adolescent drunk driving, and skin cancer.

Quantitative Applications in the Social Sciences

A SAGE UNIVERSITY PAPERS SERIES

1. Analysis of Variance, 2nd Edition *Iversen/Norpoth*
2. Operations Research Methods *Nagel/Neef*
3. Causal Modeling, 2nd Edition *Asher*
4. Tests of Significance *Henkel*
5. Cohort Analysis *Glenn*
6. Canonical Analysis and Factor Comparison *Levine*
7. Analysis of Nominal Data, 2nd Edition *Reynolds*
8. Analysis of Ordinal Data *Hildebrand/Laing/Rosenthal*
9. Time Series Analysis, 2nd Edition *Ostrom*
10. Ecological Inference *Langbein/Lichtman*
11. Multidimensional Scaling *Kruskal/Wish*
12. Analysis of Covariance *Wildt/Ahtola*
13. Introduction to Factor Analysis *Kim/Mueller*
14. Factor Analysis *Kim/Mueller*
15. Multiple Indicators *Sullivan/Feldman*
16. Exploratory Data Analysis *Hartwig/Dearing*
17. Reliability and Validity Assessment *Carmines/Zeller*
18. Analyzing Panel Data *Markus*
19. Discriminant Analysis *Klecka*
20. Log-Linear Models *Knoke/Burke*
21. Interrupted Time Series Analysis *McDowall/McCleary/Meidinger/Hay*
22. Applied Regression *Lewis-Beck*
23. Research Designs *Spector*
24. Unidimensional Scaling *McIver/Carmines*
25. Magnitude Scaling *Lodge*
26. Multiattribute Evaluation *Edwards/Newman*
27. Dynamic Modeling *Huckfeldt/Kohfeld/Likens*
28. Network Analysis *Knoke/Kuklinski*
29. Interpreting and Using Regression *Achen*
30. Test Item Bias *Osterlind*
31. Mobility Tables *Hout*
32. Measures of Association *Liebetrau*
33. Confirmatory Factor Analysis *Long*
34. Covariance Structure Models *Long*
35. Introduction to Survey Sampling *Kalton*
36. Achievement Testing *Bejar*
37. Nonrecursive Causal Models *Berry*
38. Matrix Algebra *Namboodiri*
39. Introduction to Applied Demography *Rives/Serow*
40. Microcomputer Methods for Social Scientists, 2nd Edition *Schrodt*
41. Game Theory *Zagare*
42. Using Published Data *Jacob*
43. Bayesian Statistical Inference *Iversen*
44. Cluster Analysis *Aldenderfer/Blashfield*
45. Linear Probability, Logit, and Probit Models *Aldrich/Nelson*
46. Event History Analysis *Allison*
47. Canonical Correlation Analysis *Thompson*
48. Models for Innovation Diffusion *Mahajan/Peterson*
49. Basic Content Analysis, 2nd Edition *Weber*
50. Multiple Regression in Practice *Berry/Feldman*
51. Stochastic Parameter Regression Models *Newbold/Bos*
52. Using Microcomputers in Research *Madron/Tate/Brookshire*
53. Secondary Analysis of Survey Data *Kiecolt/Nathan*
54. Multivariate Analysis of Variance *Bray/Maxwell*
55. The Logic of Causal Order *Davis*
56. Introduction to Linear Goal Programming *Ignizio*
57. Understanding Regression Analysis *Schroeder/Sjoquist/Stephan*
58. Randomized Response *Fox/Tracy*
59. Meta-Analysis *Wolf*
60. Linear Programming *Feiring*
61. Multiple Comparisons *Klockars/Sax*
62. Information Theory *Krippendorff*
63. Survey Questions *Converse/Presser*
64. Latent Class Analysis *McCutcheon*
65. Three-Way Scaling and Clustering *Arabie/Carroll/DeSarbo*
66. Q Methodology *McKeown/Thomas*
67. Analyzing Decision Making *Louviere*
68. Rasch Models for Measurement *Andrich*
69. Principal Components Analysis *Dunteman*
70. Pooled Time Series Analysis *Sayrs*
71. Analyzing Complex Survey Data *Lee/Forthofer/Lorimor*
72. Interaction Effects in Multiple Regression, 2nd Edition *Jaccard/Turrisi*
73. Understanding Significance Testing *Mohr*
74. Experimental Design and Analysis *Brown/Melamed*
75. Metric Scaling *Weller/Romney*
76. Longitudinal Research, 2nd Edition *Menard*
77. Expert Systems *Benfer/Brent/Furbee*
78. Data Theory and Dimensional Analysis *Jacoby*
79. Regression Diagnostics *Fox*
80. Computer-Assisted Interviewing *Saris*
81. Contextual Analysis *Iversen*
82. Summated Rating Scale Construction *Spector*
83. Central Tendency and Variability *Weisberg*
84. ANOVA: Repeated Measures *Girden*
85. Processing Data *Bourque/Clark*
86. Logit Modeling *DeMaris*
87. Analytic Mapping and Geographic Databases *Garson/Biggs*
88. Working With Archival Data *Elder/Pavalko/Clipp*
89. Multiple Comparison Procedures *Toothaker*
90. Nonparametric Statistics *Gibbons*
91. Nonparametric Measures of Association *Gibbons*
92. Understanding Regression Assumptions *Berry*
93. Regression With Dummy Variables *Hardy*
94. Loglinear Models With Latent Variables *Hagenaars*
95. Bootstrapping *Mooney/Duval*
96. Maximum Likelihood Estimation *Eliason*
97. Ordinal Log-Linear Models *Ishii-Kuntz*
98. Random Factors in ANOVA *Jackson/Brashers*
99. Univariate Tests for Time Series Models *Cromwell/Labys/Terraza*
100. Multivariate Tests for Time Series Models *Cromwell/Hannan/Labys/Terraza*

Other volumes in this series listed on outside back cover

Quantitative Applications in the Social Sciences

A SAGE UNIVERSITY PAPERS SERIES

101. **Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models** *Liao*
102. **Typologies and Taxonomies** *Bailey*
103. **Data Analysis: An Introduction** *Lewis-Beck*
104. **Multiple Attribute Decision Making** *Yoon/Hwang*
105. **Causal Analysis With Panel Data** *Finkel*
106. **Applied Logistic Regression Analysis, 2nd Edition** *Menard*
107. **Chaos and Catastrophe Theories** *Brown*
108. **Basic Math for Social Scientists: Concepts** *Hagle*
109. **Basic Math for Social Scientists: Problems and Solutions** *Hagle*
110. **Calculus** *Iversen*
111. **Regression Models: Censored, Sample Selected, or Truncated Data** *Breen*
112. **Tree Models of Similarity and Association** *James E. Corter*
113. **Computational Modeling** *Taber/Timpone*
114. **LISREL Approaches to Interaction Effects in Multiple Regression** *Jaccard/Wan*
115. **Analyzing Repeated Surveys** *Firebaugh*
116. **Monte Carlo Simulation** *Mooney*
117. **Statistical Graphics for Univariate and Bivariate Data** *Jacoby*
118. **Interaction Effects in Factorial Analysis of Variance** *Jaccard*
119. **Odds Ratios in the Analysis of Contingency Tables** *Rudas*
120. **Statistical Graphics for Visualizing Multivariate Data** *Jacoby*
121. **Applied Correspondence Analysis** *Clausen*
122. **Game Theory Topics** *Fink/Gates/Humes*
123. **Social Choice: Theory and Research** *Johnson*
124. **Neural Networks** *Abdi/Valentin/Edelman*
125. **Relating Statistics and Experimental Design: An Introduction** *Levin*
126. **Latent Class Scaling Analysis** *Dayton*
127. **Sorting Data: Collection and Analysis** *Coxon*
128. **Analyzing Documentary Accounts** *Hodson*
129. **Effect Size for ANOVA Designs** *Cortina/Nouri*
130. **Nonparametric Simple Regression: Smoothing Scatterplots** *Fox*
131. **Multiple and Generalized Nonparametric Regression** *Fox*
132. **Logistic Regression: A Primer** *Pampel*
133. **Translating Questionnaires and Other Research Instruments: Problems and Solutions** *Behling/Law*
134. **Generalized Linear Models: A Unified Approach** *Gill*
135. **Interaction Effects in Logistic Regression** *Jaccard*
136. **Missing Data** *Allison*
137. **Spline Regression Models** *Marsh/Cormier*
138. **Logit and Probit: Ordered and Multinomial Models** *Borooah*
139. **Correlation: Parametric and Nonparametric Measures** *Chen/Popovich*
140. **Confidence Intervals** *Smithson*

Visit our website at www.sagepub.com

ISBN 0-7619-2742-5

90000>

9 780761 927426

SAGE PUBLICATIONS
International Educational and Professional Publisher
Thousand Oaks London New Delhi