Methodology in the Social Sciences
David A. Kenny, *Series Editor*

PRINCIPLES AND PRACTICE OF STRUCTURAL EQUATION MODELING
Rex B. Kline

SPECTRAL ANALYSIS OF TIME-SERIES DATA
Rebecca M. Warner

A PRIMER ON REGRESSION ARTIFACTS
Donald T. Campbell and David A. Kenny

REGRESSION ANALYSIS FOR CATEGORICAL MODERATORS
Herman Aguinis

HOW TO CONDUCT BEHAVIORAL RESEARCH OVER THE INTERNET:
A BEGINNER'S GUIDE TO HTML AND CGI/PERL
R. Chris Fraley

Regression Analysis for Categorical Moderators

Herman Aguinis

SERIES EDITOR'S NOTE by David A. Kenny

THE GUILFORD PRESS
New York London

2004

1

What Is a Moderator Variable and Why Should We Care?

If we want to know how well we are doing in the biological, psychological, and social sciences, an index that will serve us well is how far we have advanced in our understanding of the moderator variables of our field.

—HALL AND ROSENTHAL (1991, p. 447)

The promising results obtained to date indicate that further work should be conducted in the controversial research area of moderator variables.

—CHEEK (1989, p. 281)

The advancement and theoretical sophistication of the social sciences have motivated researchers to go beyond first-order effects and understand moderated, also labeled interactive, relationships. More specifically, researchers are interested in testing whether the relationship between two variables changes depending on the value of a discrete grouping variable. For instance, social scientists have recently investigated such diverse questions as the following:

- Does the relationship between conflict with parents and depressive symptoms change based on adolescents' nationality (i.e., United States vs. China) (Greenberger, Chen, Tally, & Dong, 2000)?
- Does a firm's risk aversion affect its attractiveness to potential employees differently depending on firm ownership type (i.e., state-owned, international joint venture, or wholly owned foreign enterprise) (Turban, Lau, Ngo, Chow, & Si, 2001)?

- Does the relationship between perfectionism and bulimic symptoms in female college students change depending on perceived weight status (i.e., overweight vs. not overweight) (Vohs, Bardone, Joiner, Abramson, & Heatherton, 1999)?
- Does juggling of work and family roles affect self-reports of negative affect and calmness differently depending on the setting of the activities (i.e., work vs. home) (Williams & Alliger, 1994)?
- Does a preemployment test exhibit predictive bias such that the relationship between test scores and measures of job performance depends on ethnicity (e.g., minority or majority) and gender (i.e., male or female) (Rotundo & Sackett, 1999; Saad & Sackett, 2002; Society for Industrial and Organizational Psychology [SIOP], 1987; Te Nijenhuis & Van der Flier, 1999)?
- Does psychosis affect depression differently for various age groups (e.g., 18–39 vs. 40–59, or 60–79) (Jorm et al., 2000)?
- Does the relationship between personal goals and job performance change based on type of goal (easy vs. hard) (Tubbs, 1993)?
- Does the relationship between the strategy an importer chooses to use and the renewal of an importing contract depend on the country of origin of the importer (Peru vs. United States) (Marshall & Boush, 2001)?
- Does the relationship between proactive job search and long-term mental health among unemployed individuals change based on reemployment status (reemployed vs. unemployed) (Wanberg, 1997)?
- Do the effects of advertising content on brand attitude vary across levels of brand loyalty (e.g., high vs. low) (LeClerc & Little, 1997)?

Each of the preceding questions shares the same interest in whether the (presumably causal) relationship between two quantitative variables X and Y changes based on the value of a third discrete grouping variable Z. This third variable Z is labeled a moderator of the relationship between variables X and Y when the nature of this relationship is contingent upon Z (Stone, 1988; Stone-Romero & Liakhovitski, 2002; Zedeck, 1971). For example, a moderating-effect hypothesis could be that the effect of setting personal goals (i.e., X) on job performance (i.e., Y) depends on the level of goal difficulty (i.e., Z, easy vs. hard) such that the relationship is stronger for hard as compared to easy goals. Note that one can also describe the moderating effect as an interaction between X and Z. Furthermore, because interactive relationships are symmetrical, one could refer to the moderating effect of Z on the X–Y relationship, or to the moderating effect of X on the Z–Y relationship. Which variable is chosen as the moderator depends on the substantive research question.

The grouping moderator variable can be experimentally manipulated (e.g., goal difficulty) or naturally occurring (e.g., gender). For example, it may be the case that the "easy goal" group exhibits a strong and positive relationship between personal goals and job performance whereas the "hard goal" group shows a positive but weaker relationship. Alternatively, it could be that the hard goal group shows no goal–performance relationship whatsoever, or even a strong and negative relationship. Thus, in general, the grouping variable Z is a moderator when the X–Y relationship is not the same across the groups under consideration (e.g., easy vs. hard goal, men vs. women, United States vs. China nationals, and so forth). Note that the nature of the relationship between X and Y is such that X may cause Y (in the case of experimental designs), or X and Y may covary (in the case of nonexperimental designs). In general, however, it is typically assumed that X is a causal antecedent to Y.

Although the interest in moderated relationships has increased dramatically over the past 20 years, social scientists have noted the existence of moderator variables for almost half a century (Abelson, 1952; Edwards, 1954; Frederiksen & Melville, 1954; Gaylord & Carroll, 1948; Ghiselli, 1956; Saunders, 1955, 1956). The label "moderator variable" seems to have been used first by Saunders in 1955 but, according to Zedeck (1971), the concept had been discussed previously. For example, Court (1930) used "joint causation," Gaylord and Carroll (1948) used "population control variable," and Frederiksen and Melville (1954) used "subgrouping variable." Even after Saunders (1956) published his paper "Moderator Variables in Prediction," there were authors who continued to use terms other than that to describe moderating effects. Ghiselli (1956, 1960a, 1960b) used "predictability variable," Toops (1959) used "referent variable," Grooms and Endler (1960) used "modifier variable," Johnson (1960) used "homologizer variable," and Velicer (1972) used "heterogeneous regression." At present, although other terms are used occasionally (Sharma, Durand, & Gur-Arie, 1981), moderator appears to have become the accepted term in most social science disciplines.

WHY SHOULD WE STUDY MODERATOR VARIABLES?

Moderator variables are playing an increasingly important role in social science research and practice. On the one hand, researchers may look for moderators to ascertain whether a causal law is general. In situations where researchers seek to find generalizability, the ideal outcome is the finding that there are no moderators (Hunter, Schmidt, & Rauschenberger, 1984). On the other hand, researchers may look for moderators in an attempt to improve the fit of their models, given that main effects alone may not provide sufficient accuracy in prediction. In these situations, the ideal outcome is the finding that there are strong moderated relationships. Regardless of the outcome, the study of moderator variables has implications for both theory and practice because it provides information on the boundary conditions for the relationships of interest.

Consider the situation in which a researcher finds a positive and very strong effect between preemployment test scores and job performance in a sample of men and a weak and almost nonexistent relationship in a sample of women (cf. Stricker, Rock, & Burton, 1993; Young, 1994). By not identifying the moderating effect of gender and considering the overall test scores–performance relationship only, the researcher would conclude that there is a moderate relationship between test scores and performance (i.e., the average test scores–performance relationship across the two groups). Similarly, the British writer and politician Benjamin Disraeli (1804–1881) noted the following (Huff, 1954): "A man eats a loaf of bread, and another man eats nothing; statistics is the science that tells us that each of these men ate half a loaf of bread." In the aforementioned preemployment testing situation, a similarly incorrect conclusion is that there is a moderate relationship between test scores and performance, whereas in actuality the relationship is positive and very strong for men and virtually nonexistent for women.

From a theory point of view, the erroneous conclusion about the absence of a moderating effect of gender precludes the researcher from understanding the sources of the differential relationship across groups (Baker & Yardley, 2002). For example, are testing procedures implemented in a discriminatory way? Is one gender-based sample more homogeneous than the other? Are there meaningful psychological differences in how men and women interpret the test in question? Failure to understand these differential relationships prevents researchers from learning about boundary conditions for the causal effects in question, and therefore is likely to delay the advancement of theory.

From a practice point of view, using this particular test for personnel selection purposes is likely to lead to errors in prediction (underprediction of future performance for men and overprediction of future performance for women). In turn, these errors in prediction are likely to have detrimental effects for organizational productivity as well as employee satisfaction and well-being, not to mention potential litiga-

tion. In short, not understanding the moderating effect has important implications for both theory and practice.

DISTINCTION BETWEEN MODERATOR AND MEDIATOR VARIABLES

It is important to distinguish a moderating from a mediating effect. These effects are distinct and are not necessarily mutually exclusive (e.g., Sheeran & Abraham, 2003). A variable is a mediator of an X to Y relationship when it accounts for the causal relation between X and Y (Baron & Kenny, 1986). Mediators are also called "intervening" or "process" variables because they explain the relationship between two variables. In other words, mediators give us information on *why* or *by what mechanism* X causes Y (Frone, 1999). In contrast, a moderator variable explains changes in the nature of the X to Y effect. That is, a moderator explains *when* or *under what conditions* X causes Y (Frone, 1999). As noted by Baron and Kenny (1986), "whereas moderator variables specify when certain effects will hold, mediators speak to how or why such effects occur" (p. 1176). Figure 1.1 shows these relationships graphically. The top panel shows a moderated relationship and the bottom panel shows a mediated relationship.

As an example, consider the literature on work stress and alcohol use (Frone, 1999; see Sheeran and Abraham, 2003, for an example in the social psychological literature). Researchers are interested in investigating

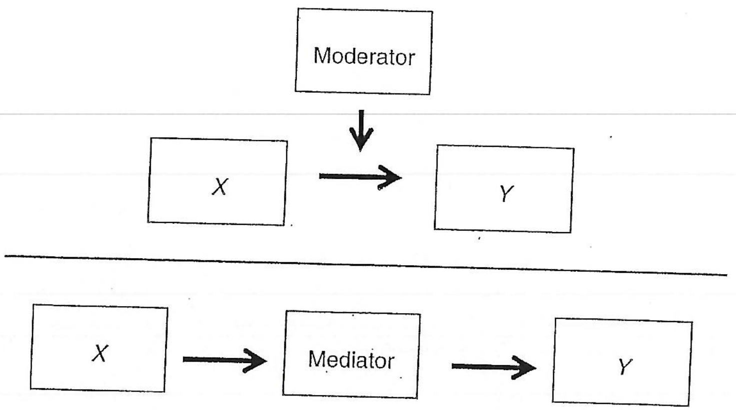


FIGURE 1.1. Representation of a variable serving as a moderator (top) and mediator (bottom).

the effect of work stressors (i.e., aversive work conditions) on employee alcohol use. Some work stressors commonly investigated include danger- ous work conditions, noxious physical work environments (e.g., noise and dirt), interpersonal conflict with supervisors or coworkers, and un- fair treatment regarding pay (Frone, 1999). Researchers have tested mod- erated relationships in an effort to understand which are the groups of em- ployees for whom the effect of work stressors on alcohol use is stronger. For instance, work stressors were found to have an effect on alcohol use only for those employees whose work role is psychologically important to their self-definition. On the other hand, for employees whose work is not psychologically important, there was no effect of work stressors on alcohol use (Frone, Russell, & Cooper, 1997). In addition to having an in- terest in moderated relationships, researchers have also investigated hypothesized mediated relationships in an effort to understand the un- derlying mechanisms responsible for the effect of work stressors on alcohol use. For example, anxiety was found to mediate the relationship between the work stressor "poor relationship with supervisors and coworkers" and average weekly alcohol consumption (Vasse, Nijhuis, & Kok, 1998). The top part of Figure 1.2 graphically displays the moder- ating effect of work role centrality (i.e., high vs. low), and the bottom dis- plays the mediating effect of anxiety. This book focuses on moderating effects. Readers interested in mediation should consult Baron and Kenny (1986), James and Brett (1984), Judd, Kenny, and McClelland (2001), and Shrout and Bolger (2002).

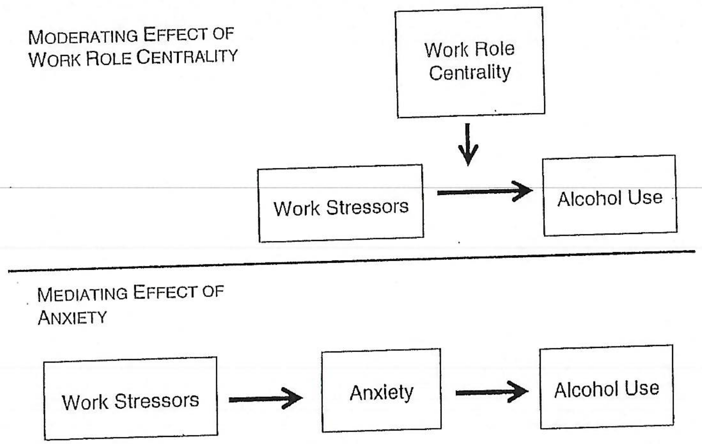


FIGURE 1.2. Representation of work role centrality as a moderator variable (top) and anxiety as a mediator variable (bottom).

IMPORTANCE OF A PRIORI RATIONALE IN INVESTIGATING MODERATING EFFECTS

In many cases, researchers have specific theory-based predictions to guide hypotheses about potential moderator variables (Chaplin, 1997). For example, Greenberger et al. (2000) hypothesized that nationality (i.e., United States vs. China) would moderate the link between the quality of family relationships and depressed mood such that the rela- tionship would be stronger for a Chinese sample as compared to a U.S. sample. This prediction was based on the importance of harmonious re- lationships, especially within one's family, in Chinese society (Aguinis & Roth, 2003). Because of the high importance placed on harmonious family relationships, there was a strong theory-based rationale to pre- dict that the link between quality of family relationships and depressed mood would be stronger for Chinese as compared to U.S. participants.

In other cases, researchers routinely test for the moderating effects of such variables as gender and ethnicity because tests of that kind are recommended as best practices in specific research domains (Bartlett, Bobko, Mosier, & Hannan, 1978; Bartlett & O'Leary, 1969). For example, over 30 years ago, Einhorn and Bass (1971) issued the recommendation that "it is *always* necessary to investigate the possibility of different regression functions for different groups, since otherwise one might be guilty of discrimination in the use of tests" (p. 263). More recently, the *Principles for the Validation and Use of Personnel Selection Procedures* (SIOP, 1987) recommended such investigations to gather evidence regarding a test's fairness for various groups (Jones, 1973). Thus, al- though there may not be a specific theory-related rationale for the moder- ator test, professional practice dictates that this analysis be conducted.

In yet a third type of situation, however, researchers choose to ex- amine the potential moderating effects of grouping variables such as gender and ethnicity without a strong rationale for such investigations. Given that the data have been collected and that conducting a modera- tor analysis is just a few mouse clicks away, many researchers are tempted to conduct such exploratory analysis in the absence of any type of hypothesis or justification. It may even be the case that a researcher finds an unexpected moderating effect. Although it is permissible to conduct such exploratory examinations, results should be interpreted with great caution and should be replicated using independent samples. As described in more detail in Chapter 5, unless certain conditions ex- ist, it is likely that moderating effects will go undetected in a sample even if they do exist in the population. Moderator variables are difficult to detect even when the moderator test is the focal issue in a research study and a researcher has designed the study specifically with the moderator test in mind. If a moderator test is conducted without a clear a priori justification, based on theory or recommended professional practice, it is likely that the research design will not be conducive to assessing the moderating effect accurately. In short, an a priori rationale should guide the investigation of moderator variables. Otherwise, the conclusions regarding moderating effects may be erroneous.

CONCLUSIONS

- This chapter defined the concept of a moderator variable, illustrated the pervasiveness and importance of moderated relationships in several social sciences, distinguished moderator from mediator variables, and discussed the importance of having good a priori justification for studying moderated relationships.
- If the relationship between two variables is not the same across the various groups under consideration, the grouping variable is a moderator (i.e., there is an interaction effect). Social scientists must have a good understanding of the moderator variables in their fields. Otherwise it is likely that resources will not be used wisely and interventions will lead to unintended, and even opposite, outcomes. For instance, does the implementation of a performance appraisal system based on individual goals lead to the same level of performance improvement across cultures? Not having a good understanding of this issue may lead corporate headquarters to implement the same system in all subsidiaries worldwide. Although positive outcomes may be the result in an individualistic culture like the United States, negative outcomes—including decreased productivity, job dissatisfaction, and a deterioration of the level of trust with headquarters—are possible in a collectivistic culture like China (Aguinis & Roth, 2003).
- Although the focus of this book is on moderator variables, researchers must be aware of the distinction between moderator and mediator variables. Moderator variables provide information regarding the conditions under which an effect or relationship is likely to be stronger. Mediator variables provide information regarding the mechanisms likely to be responsible for the effect or relationship in question. Moderators and mediators are not mutually exclusive, and a hypothesized model may include both (e.g., Frone, 1999).
- This chapter has also emphasized the importance of having good justification before launching into the analysis of moderator variables. Because analysis of moderators is like a minefield, filled with difficulties that are typically unknown to the researcher conducting the analysis, going on a "fishing expedition" in search of moderators is likely to lead to conclusions that are not replicable in subsequent studies.

- Chapter 2 describes how to conduct a moderator analysis using moderated multiple regression (MMR) in a nontechnical way. It also describes recently published research in several social sciences that shows how pervasive MMR is for testing whether a grouping variable moderates the relationship between two quantitative variables.

2

Moderated Multiple Regression

It now seems clear that moderated regression analysis is the appropriate inferential procedure when the underlying theory postulates differences in the form of a relationship between two variables as a function of some moderator variable.
—CHAMPOUX AND PETERS (1987, p. 252)

The previous chapter defined the concept of a moderator variable, provided illustrations of moderators in the social sciences, showed the difference between moderators and mediators, and discussed the importance of having good a priori justification for studying moderated relationships. This chapter describes how to test for the presence of a moderator variable using moderated multiple regression (MMR). The chapter includes a description of the MMR procedure, discusses the appropriateness of using MMR to assess moderating effects, and describes the pervasive use of MMR to determine whether a moderator variable exists. Chapter 3 then provides a step-by-step demonstration of how to implement MMR using computer programs such as SPSS.

WHAT IS MMR?

MMR is an inferential procedure which consists of comparing two different least-squares regression equations (Aiken & West, 1991; Cohen & Cohen, 1983; Jaccard, Turrisi, & Wan, 1990). An inferential procedure would not be needed if a researcher had access to the entire population of true scores (i.e., free of measurement errors). Obviously, this is

a rare event in the social sciences. Thus, researchers collect sample data and then make inferences about populations. MMR allows researchers to make the inference of whether a moderating effect is present in the population based on sample data.

Given that a researcher collects data regarding a quantitative criterion or dependent variable Y, a predictor X, and a second binary predictor Z hypothesized to be a moderator, Equation 2.1 shows the ordinary least-squares (OLS) regression equation that tests a model predicting Y from the first-order effects of X and Z:

$$Y = a + b_1X + b_2Z + e \qquad (2.1)$$

where a is the least-squares estimate of the intercept, b_1 is the least-squares estimate of the population regression coefficient for X, b_2 is the least-squares estimate of the population regression coefficient for Z, and e is a residual term. Note that the criterion Y is a quantitative variable; other procedures such as logistic regression can be used in situations where Y is categorical (Ganzach, Saporta, & Weber, 2000; Jaccard, 2001). Also, Equation 2.1 assumes the prototypical situation where the moderator is binary (i.e., has two categories). Chapter 8 addresses more complex MMR models, including those with moderators with three categories. Equation 2.1 is represented graphically in Figure 2.1.

In Equation 2.1, the regression coefficient b_1 is interpreted as the number of units that Y is predicted to increase with a 1-unit increase in X given that Z is held constant. For example, assume that Y is a measure of job performance (supervisory ratings on a 1 = *below expectations* to 7 = *above expectations* scale) and X is a preemployment test of general cognitive abilities (intelligence measured on scale ranging from 1 to 10). Given this illustrative situation, $b_1 = 2$ is interpreted as follows: For

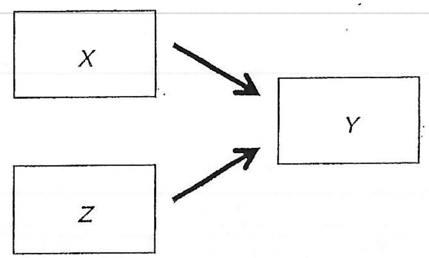


FIGURE 2.1. Graphic representation of Equation 2.1, showing the model including the first-order effects for predicting Y from X and Z.

every 1-point increase in general cognitive abilities, supervisory ratings of performance are predicted to increase 2 points (holding Z constant). As a second illustration, assume now that X is a measure of absenteeism (e.g., number of unexcused absences from work per year). A regression coefficient $b_1 = -.5$ is interpreted as follows: For every unexcused absence from work, supervisory ratings of performance are predicted to decrease half a point (holding Z constant). The interpretation of the intercept a and the regression coefficient b_2 is dictated by the coding scheme that has been used for Z. More detailed information on the interpretation of a and b_2 is provided in Chapters 3 and 8.

The multiple regression model shown in Equation 2.1 assumes that all variables identified by the theory are included in the model and that the variables are properly measured. In addition, it is assumed that the population data have the following characteristics:

1. The relationship between each of the predictors and the criterion is linear.
2. Residuals (i.e., difference between predicted and actual Y scores) exhibit homoscedasticity (i.e., constant variance across values of each predictor; that is, residuals are evenly distributed throughout the regression line).
3. Residuals are independent (i.e., there is no relationship among residuals for any subset of cases in the sample).
4. Residuals are normally distributed.
5. There is less than complete multicollinearity (i.e., perfect correlation between the predictors).

These are the usual assumptions of all ordinary least-squares (OLS) multiple regression models and are described in detail in regression textbooks (e.g., Cohen & Cohen, 1983; Cohen, Cohen, West, & Aiken, 2003; Pedhazur, 1982).

The second equation, called the MMR model, is formed by creating a new variable, the product between the predictors (i.e., $X \cdot Z$), and including it as a third term in the regression. The addition of the product term to the equation yields the following model:

$$Y = a + b_1 X + b_2 Z + b_3 X \cdot Z + e \qquad (2.2)$$

where b_3 is the sample-based least-squares estimate of the population regression coefficient for the product term. This model is represented graphically in Figure 2.2.

Consider the situation where Y is salary after graduation from college, X is number of job offers received, and Z is a hypothesized binary

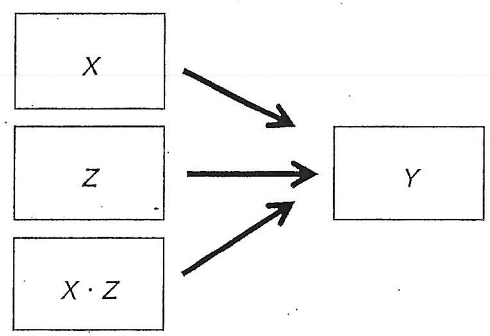


FIGURE 2.2. Graphic representation of Equation 2.2, showing the model including the first-order effects for predicting Y from X, Z, and the product term $X \cdot Z$ which carries information regarding the moderating effect of Z. This graph is conceptually equivalent to Figure 1.1 (top).

grouping moderator variable called type of industry (i.e., for-profit vs. not-for-profit). In other words, the moderator Z includes two categories or levels, and Z can be coded using one number for members of the for-profit group (i.e., 0) and a different number for members of the not-for-profit group (i.e., 1) (Chapter 8 discusses various coding schemes in detail, including situations when the moderator variable has more than two levels). A researcher forms the hypothesis that the relationship between number of job offers and starting salary is moderated by type of industry, such that the relationship is stronger in for-profit as compared to not-for-profit organizations. The hypothesis is based on the premise that because of the salary compression and fewer resources in most not-for-profit organizations, receiving numerous job offers is not likely to lead to higher starting salaries as compared to receiving fewer offers. On the other hand, for-profit organizations are expected to be able to match offers from other organizations for well-qualified applicants, which is likely to lead to higher starting salaries.

Figure 2.3 shows a scatter plot for the effect of job offers on salary for graduates applying to profit and not-for-profit organizations. This figure seems to provide support for the following hypothesis: There is a stronger relationship for the graduates aspiring to work in for-profit as compared to not-for-profit organizations. Having a larger number of job offers in hand yields greater predicted increases in salary for the graduates applying to work in for-profit as compared to not-for-profit organizations.

To formally test for the moderating effect of type of industry, a

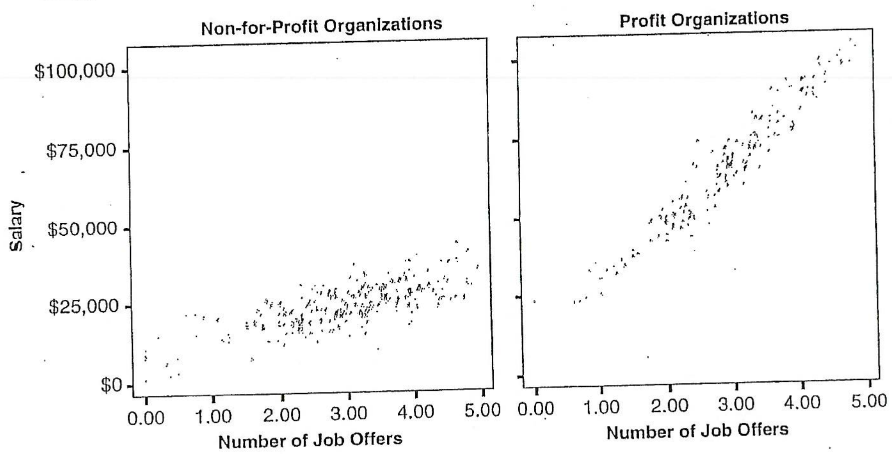


FIGURE 2.3. Hypothetical scatter plots for the relationship between number of job offers and salary for graduates applying to working for-profit (left) and not-for-profit (right) organizations.

t statistic can be computed to test the null hypothesis $\beta_3 = 0$. The term β_3 is used to symbolize the regression coefficient for the product term in the population (note that b_3 shown in Equation 2.2 is the product term coefficient in the sample) and should not be confused with beta (i.e., the standardized regression coefficient printed in most computer outputs). Conceptually, this null hypothesis tests whether the amount of change in the slope of the regression of Y on X that results from a unit change in variable Z is greater than would be expected by chance alone.

Alternatively, and equivalently, the coefficients of determination (i.e., squared multiple correlation coefficients, R^2) are compared for Equation 2.1 (i.e., R_1^2) and Equation 2.2 (i.e., R_2^2). The null hypothesis tested is H_0: $\psi_2^2 - \psi_1^2 = 0$, which can also be written as $\Delta\psi^2 = 0$. Conceptually, this null hypothesis tests whether the addition of the product term to the regression equation improves the proportion of explained variance in Y. In other words, this hypothesis answers the question of whether the moderating effect of Z helps improve the prediction of Y above and beyond the first-order effects of X and Z. Note that in some cases R_2^2 may not be statistically significant (Bedeian & Mossholder, 1994). However, the focus is whether the addition of the product term improves the fit of the model predicting Y.

To test H_0: $\psi_2^2 - \psi_1^2 = 0$, an F statistic (distributed with $k_2 - k_1$ and $N - k_2 - 1$ degrees of freedom) is computed using the following formula:

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)} \tag{2.3}$$

where k_2 is the number of predictors in Equation 2.2, k_1 is the number of predictors in Equation 2.1, and N is the total sample size. Note that the statistical significance levels (i.e., p values) associated with the t and F tests are identical (Cohen & Cohen, 1983).

As a measure of moderating effect size, most researchers choose to focus on ΔR^2, as opposed to b_3, though it is not ideal. The reason for this choice seems to be that ΔR^2 refers to proportion of variance explained and, therefore, is a common metric that can be used to compare effect sizes across studies and areas of research. On the other hand, b_3 is metric specific and is referenced to the specific scales used to measure X, Y, and Z. Consequently, it is difficult to use b_3 to assess the relative size of moderating effects across studies using different scales, even if the studies tested the same moderating-effect hypothesis. A more detailed discussion of ways to assess effect size and the practical importance of moderating effects is provided in Chapter 9.

Furthermore, referring back to Equation 2.2, one may wish to compute a t statistic to test the null hypothesis $\beta_2 = 0$. Note that the term β_2 denotes the regression coefficient for the grouping variable Z in the population, and b_2 is its sample-based least-squares estimate. One may also wish to interpret the meaning of the intercept a. The interpretation of a and b_2 varies, depending on which coding scheme is used. More information on the interpretation of the intercept and coefficients for first-order effects in the presence of an interaction is provided in Chapter 3.

Although no formal tests are performed at this point (a detailed analysis of a fully worked-out example is provided in Chapter 3), Figure 2.3 shows that the slopes seem to differ across groups. Specifically, as hypothesized, there is a steeper (i.e., stronger) slope for the for-profit as compared to the not-for-profit group. Of course, this tentative conclusion based on "eyeballing the data" should be tested analytically by examining the statistical significance of b_3. In addition, Figure 2.3 shows that the intercepts also seem to differ across the groups. Specifically, the regression line crosses the Y-axis at a value of about \$9,600 for the for-profit group and at a value of about \$8,700 for the not-for-profit group. Again, the conclusion regarding differences between intercepts is tentative until a formal test is conducted. Chapter 3 provides a detailed demonstration of how to implement an MMR analysis using a widely available commercial computer package.

ENDORSEMENT OF MMR AS AN APPROPRIATE TECHNIQUE

MMR has been recognized as an appropriate technique to assess the presence of moderator variables for half a century (Saunders, 1955, 1956). In spite of the availability of MMR, in the past three decades or so a number of alternative methods have been proposed (Anderson, Stone-Romero, & Tisak, 1996; Arnold, 1982, 1984; Blood & Mullet, 1977; Bobko, 1986; Darrow & Kahl, 1982; Kahl & Darrow, 1984; Morris, Sherman, & Mansfield, 1986). For example, Darrow and Kahl (1982) proposed that the product term be entered before the X and Z predictors in Equation 2.2, based on the argument that this procedure enhances the probability of detecting a moderating effect. Bobko (1986) suggested a set of planned comparisons driven by theory, based on the argument that this procedure increases the power of the moderator test, preserves overall degrees of freedom, and reduces the experiment-wide error rate. Anderson et al. (1996) proposed the use of errors-in-variables regression (EIVR) in lieu of MMR based on the argument that EIVR may not be as adversely affected by measurement error as MMR, thus providing less biased population estimates for the regression coefficients. Hunter and Schmidt (1978) suggested the comparison of X–Y correlation coefficients across groups as a test of the moderating effect. And Morris et al. (1986) proposed the use of principal-components regression (PCR) as an alternative to MMR, based on the argument that PCR is not as adversely affected by multicollinearity (i.e., the correlation between X and Z) as MMR.

Although some of the methods proposed as alternatives to MMR received favorable reviews initially, with the passage of time these methods have been shown to be problematic, and in some cases inappropriate. Stone (1988) and Stone and Hollenbeck (1984, 1989) critically discussed the methods proposed by Arnold (1982, 1984) and Darrow and Kahl (1982). They concluded that MMR was superior, based on theoretical and statistical considerations. Furthermore, the method recommended by Darrow and Kahl (1982; see also Kahl & Darrow, 1984) violates basic principles of multiple regression because the first-order effects must precede (or be entered simultaneously with) the product term in the regression equation; this "backward regression" technique (Stone, 1988) was criticized on the basis of logical and methodological arguments (Stone, 1986; Stone & Hollenbeck, 1984; Tisak, 1994; Wise, Peters, & O'Connor, 1984). The techniques suggested by Bobko (1986) and Morris et al. (1986) were scrutinized by the late Lee Cronbach (1987), who concluded that they are "problematic" and "unacceptable," respectively. Echoing Cronbach's assessment, Stone (1988) criticized Bobko's (1986) proposed technique, arguing that although comparing a particular cell mean to the mean of the other three cell means in the context of a 2×2 design provides information about a main effect contrast, it does not provide direct evidence regarding the moderating effect. In addition, Dunlap and Kemery (1987) showed that Morris et al.'s (1986) finding of a nonsignificant moderating effect with MMR and significant effects using PCR may have been a result of an artifact of PCR. Hunter and Schmidt's (1978) proposed method to compare correlation coefficients across groups (i.e., "subgroup analysis") considers the correlation coefficients only and does not include a consideration of prediction of criterion scores across moderator-based subgroups. In other words, comparing correlation coefficients (i.e., differential validity) only detracts attention from the more global issue of differential prediction (Bobko & Bartlett, 1978). Instead, using MMR involves a regression equation including means, standard deviations, *and* correlation coefficients (Bobko & Bartlett, 1978). Thus, although they have been treated interchangeably (e.g., Boehm, 1977), differential validity and differential prediction are two related yet distinct issues (Drasgow & Kang, 1984). Finally, Anderson et al.'s (1996) simulation demonstrated that EIVR's parameter estimates were superior to MMR's estimates when both sample size and reliabilities of the predictors were high, but MMR outperformed EIVR in the more common social science situations where reliabilities or sample size are low. In summary, several independent evaluations conducted over the past four decades indicate that MMR is an appropriate method for assessing the effects of moderator variables (Aiken & West, 1991; Cleary, 1968; Cohen & Cohen, 1983; Einhorn & Bass, 1971; Evans, 1991a, 1991b; Fisicaro & Tisak, 1994; Friedrich, 1982; Jaccard et al., 1990; Saunders, 1956; Stone, 1988; Stone & Hollenbeck, 1984, 1989; Stone-Romero & Anderson, 1994; Zedeck, 1971).

Perhaps as a consequence of the numerous independent evaluations concluding that MMR is an appropriate technique for estimating moderating effects, reports issued by several professional organizations of measurement scholars and practitioners have also provided an endorsement for the use of MMR (Sackett & Wilk, 1994). More specifically, such endorsements are found in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999) and the *Principles for the Validation and Use of Personnel Selection Procedures* (SIOP, 1987). In fact, the latest edition of the *Standards* includes the following statement endorsing the use of MMR (Fairness in Testing and Test Use, Standard 7.6):

> When empirical studies of differential prediction of a criterion for members of different subgroups are conducted, they should include regression equations (or an appropriate equivalent) computed separately for each group or treatment under consideration or an analysis in which the group or treatment variables are entered as moderator variables. (p. 82)

In general, "Under one broadly accepted definition, no bias exists if the regression equations relating the test and the criterion are indistinguishable for the groups in question" (AERA, APA, & NCME, 1999, p. 79).

In sum, numerous investigations have concluded that MMR is an appropriate technique to assess the effects of categorical moderator variables, and MMR has been endorsed by several professional organizations. Consequently, it is rarely surprising that MMR is used so pervasively in the social sciences. But how frequently is MMR used in published research? This question is answered next.

PERVASIVE USE OF MMR IN THE SOCIAL SCIENCES: LITERATURE REVIEW

In some research areas, MMR is clearly the dominant, and often exclusive, methodological tool for assessing moderating effects. One example is the management accounting literature that focuses on the use of budgets in organizations. In this research area, the key substantive question is whether the relationship between the use of budgets and a number of outcome variables (e.g., individual behavioral effects) is contingent on various moderating organizational variables (Hartmann & Moers, 1999). MMR is used so frequently that it "has become the dominant statistical technique in budgetary research for testing contingency hypotheses" (Hartmann & Moers, 1999, p. 292). A second example is the clinical psychology literature that focuses on aptitude by treatment interactions resulting in various psychotherapy outcomes (Smith & Sechrest, 1991). In this literature, the substantive question is whether the relationship between psychotherapy (i.e., treatment) and psychopathology is contingent on personal characteristics of patients (i.e., aptitude). MMR is also a data analysis tool of choice in this research domain (Smith & Sechrest, 1991). Yet a third illustration of an entire research area for which MMR is the technique of choice is the job design literature (Champoux & Peters, 1980). In general, job design researchers examine whether the relationship between job design (e.g., adding task variety) and various outcomes (e.g., job satisfaction and work motivation) is moderated by individual and organizational characteristics (e.g., individuals with strong vs. weak growth needs) (Hackman & Oldham, 1976). Champoux and Peters (1980) reviewed 10 years of job design literature and concluded that MMR is the preferred data analysis technique in this research area. Finally, the vast majority of textbooks describing analysis of covariance (ANCOVA) recommend that researchers first check whether the relationship between the criterion and the covariate (i.e., control variable) is similar across groups (e.g., Cohen et al., 2003, pp. 350–351). Thus, before conducting an ANCOVA, researchers are advised to use MMR to test if this important assumption is tenable. Consequently, every researcher who is aware of this critical assumption is likely to use MMR before conducting an ANCOVA.

The above shows how pervasive the use of MMR is in several social sciences. However, in an attempt to quantify the frequency of use of MMR, Aguinis, Beaty, Boik, and Pierce (2003) reviewed all issues of the *Journal of Applied Psychology* (JAP), *Personnel Psychology* (PP), and *Academy of Management Journal* (AMJ) from 1969 to 1998 and identified all articles reporting a discrete moderator analysis test using MMR. Although the review included only three journals, these journals are among the most influential publications devoted to empirical research in applied psychology and management (Starbuck & Mezias, 1996). Thus, they serve as good illustrations of top-tier social science journals devoted to publishing original empirical work.

The criteria for counting a study as using MMR included the following:

1. At least one MMR analysis was included as part of the study.
2. The MMR analysis included a quantitative criterion Y.
3. The MMR analysis included a quantitative predictor X.
4. The MMR analysis included a categorical moderator Z (cf. Equation 2.2).

Aguinis et al. (2003) found 106 articles that fit the preceding criteria. Typically, researchers conducted more than one MMR analysis in any given study. Thus, the total number of reported MMR analyses was 636. Figure 2.4 shows the distribution of MMR analysis over the 30-year period (i.e., 1969–1998) included in the review. As can be seen in this graph, the first article using MMR to assess effects of categorical moderator variables was published in 1977. Also, there is an upward trend in the use of MMR over time. Overall, the frequency of MMR use remained at a high level of approximately 20–40 analyses per year since the mid-1980s.

Given these results based on an admittedly selected set of three

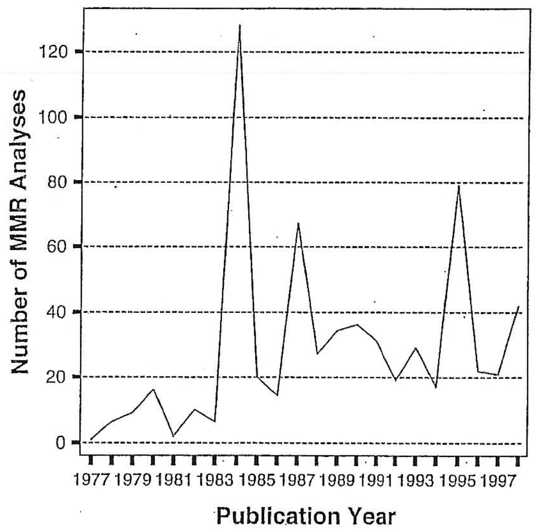


FIGURE 2.4. Number of MMR analyses of categorical moderator variables reported in *Academy of Management Journal, Journal of Applied Psychology,* and *Personnel Psychology* between January 1969 and December 1998. Frequencies for 1969–1976 = 0. Adapted from Aguinis, Beaty, Boik, and Pierce (2003).

journals, one can only guess that the number of MMR analyses reported in all social sciences journals annually is probably in the hundreds. Extrapolating from the Aguinis et al. (2003) review, the number is increasing over time.

CONCLUSIONS

- This chapter described the MMR model and its basic statistical assumptions. MMR is an extension of a multiple regression equation that includes an additional predictor carrying information regarding the moderating effect. The test of the moderating effect consists of assessing whether the regression coefficient associated with the product term is different from zero in the population; this test is conducted by computing a t statistic. Alternatively, and equivalently, one can assess whether the inclusion of the product term in the regression equation improves our ability to predict the criterion; this test is conducted by computing an F statistic. The statistical significance level of both tests is identical and, therefore, the decision re-

garding the rejection of the null hypothesis of no moderating effect in the population is identical regardless of which statistical test is used. An examination of the regression coefficient associated with the product term gives us information on whether the Y on X slope differs across moderator-based subgroups.

- MMR is an appropriate method for estimating moderating effects of categorical variables. Because of the endorsement of using MMR, the procedure is used pervasively in the social sciences. Extrapolating from a selective review of just three journals leads to the conclusion that the number of MMR analyses including categorical variables published annually in social science journals is in the hundreds. This review illustrates that MMR is a method of choice for estimating moderated relationships including categorical moderator variables.

- The next chapter provides a step-by-step demonstration of how to conduct an MMR analysis using computer programs with an actual illustration using SPSS.

3

Performing and Interpreting Moderated Multiple Regression Analyses Using Computer Programs

I do not fear computers. I fear lack of them.
—ISAAC ASIMOV

Computing is not about computers any more. It is about living.
—NICHOLAS NEGROPONTE

The previous two chapters described the importance of moderator variables for theory and practice, the MMR model, and the pervasive use of MMR in the social sciences. This chapter describes how to conduct an MMR analysis using computer programs and how to interpret the resulting output. Computer programs that conduct MMR analysis have been available for over 30 years (Rock, Barone, & Linn, 1967). This chapter describes the general logic and steps that can be applied in using the commercially available packages (e.g., SAS, SPSS). However, given the widespread use of SPSS, the illustrations in this chapter use SPSS (SPSS, Inc., 1999). This consists of a research scenario involving a moderator variable hypothesis and provides a step-by-step demonstration using a data set available on the Web. To make the best use of this chapter, it would be best to adopt a hands-on approach and read the chapter text while at the same time performing each of the steps on a computer.

RESEARCH SCENARIO

Given the increased calls for accountability in higher education institutions, many universities are implementing systematic and rigorous faculty performance evaluation systems that include measures of teaching, research, and service. Typically, these three measures are combined into an overall performance score that is tied to a specific annual salary increase. Unfortunately, there are budgetary constraints that lead to salary increases that are typically low. Nevertheless, there is an effort to create a more explicit link between faculty performance and monetary rewards.

Consider the situation where we are investigating a newly implemented faculty performance evaluation system. The system was implemented last year, and the research question is whether the relationship between a faculty member's overall performance score (i.e., a combination of research, teaching, and service scores) is a good predictor of his or her salary increase. That is, given the implementation of the new system, is there a clear relationship between faculty performance and pay raises? Also, and more important, we are interested in investigating whether the relationship between performance and salary increase is moderated by tenure status (i.e., untenured vs. tenured). Because of state regulations, the overall performance score is not the only variable that plays a role in affecting salary increases. The deans of the various colleges have the authority to allocate increases in cases where specific individual salaries are substantially below market. Also, state regulations mandate that salary increases be allocated as a percentage of each faculty member's current base salary. Someone with a perfect performance score may have a 5% increase, and someone with an average performance score may have a 3% increase, but these increases are based on the current base salary. Therefore, the salary increases will be larger for the faculty member with the 5% increase, assuming both base salaries are identical. However, a 5% increase may be a smaller dollar figure than a 3% increase if the faculty member with the 3% increase has a higher base salary. Typically, tenured faculty (i.e., associate and full professors) receive higher salaries as compared to untenured faculty (i.e., assistant professors). These discrepancies can be quite large. For example, the 2000–2001 mean annual salary for associate and full professors in accredited business colleges in the United States was $87,250, whereas the mean salary for assistant professors was $73,200 (Association to Advance Collegiate Schools of Business, 2001). Similar differences exist in other social science fields such as psychology (Wicherski, Pate, & Kohout, 2001).

From a theory and past research perspective, a merit-based pay system increases motivation and is likely to lead to performance improvement if there is a direct link between performance and rewards (Kerr, 1975). However, a merit-based system can lead to perceptions of unfairness and a decrease in motivation and job satisfaction if individuals perceive that this relationship is stronger for one group (e.g., tenured faculty) as compared to another group (e.g., untenured faculty). Given that the base salary is higher for tenured faculty, we expect that tenure status will moderate the relationship between performance and salary increase such that the relationship will be stronger for the tenured group. If such a moderated relationship is found, we would recommend a change in the system.

DATA SET

The data set was generated using the program MULTIVAR (Aguinis, 1994). This program allows for the generation of up to 10 correlated and normally distributed variables, and its executable and source code versions can be downloaded from *http://carbon.cudenver.edu/~haguinis/mmr/*.

The data set includes 400 cases (i.e., faculty members) and can also be downloaded from *http://carbon.cudenver.edu/~haguinis/mmr/*. The file includes the following three variables:

- Perf: Overall performance score ranging from 1 = *unsatisfactory* to 5 = *exceeds expectations*.
- Salary: Annual salary increase measured in dollars ranging from \$13.72 to \$2,148.91.
- Tenure: Tenure status, where 0 = *tenured* and 1 = *untenured*.

Chapter 8 discusses various coding systems for the categorical moderator variable in detail; for the purpose of this illustration dummy coding is used, such that members of one of the groups are arbitrarily assigned a 0 and members of the other group are assigned a 1. This coding scheme is recommended for situations involving binary moderators because of its simplicity and ease of interpretation of the results.

In short, *Perf* is the predictor, *Salary* is the criterion, and *Tenure* is the hypothesized moderator. Figure 3.1 shows the SPSS data screen.

It is useful to first obtain descriptive statistics to understand the data set better. All statistics computer software packages include procedures to implement these calculations. In SPSS, from the main data screen click on the "Analyze" pull-down menu and choose the "Descriptive Statistics" and "Frequencies" submenus. Double click on each of the three variables to obtain descriptive information on all of them.

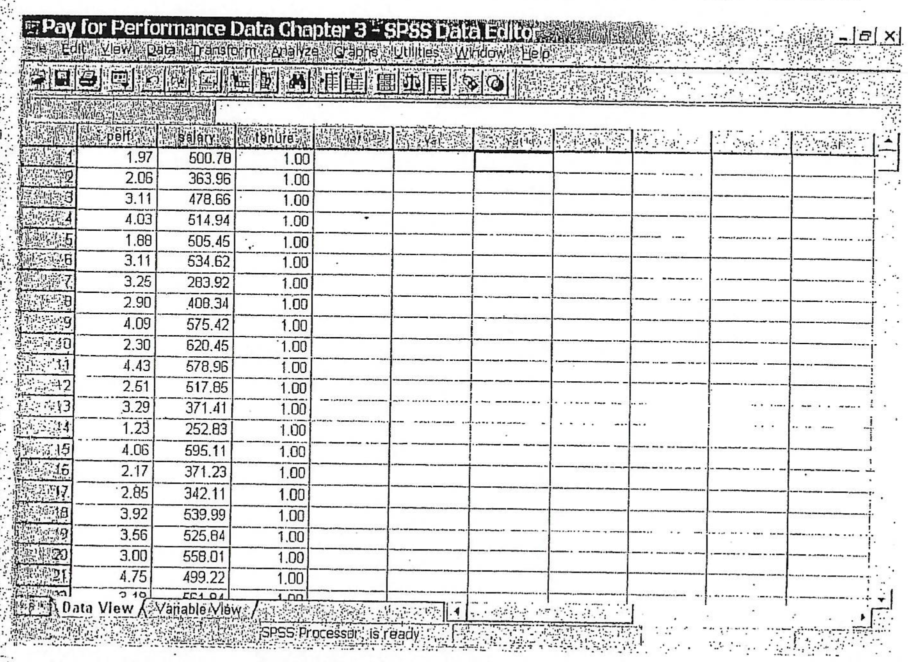


FIGURE 3.1. SPSS data screen for the Pay for Performance data set.

Figure 3.2 (top) shows the resulting screen. Then, click on the "Statistics" option and click on all "Dispersion," "Central Tendency" (except for "Sum," which is not really needed), and "Distribution" choices. Figure 3.2 (bottom) shows the resulting screen.

It is also helpful to obtain some graphs to get a better feel for the data (Tukey, 1977). Thus, from the Frequencies screen, we can click on "Charts" and choose to obtain "Histograms" and also overlap a normal curve (Figure 3.3 shows the resulting screen). Then, we can click on "Continue" and then on "OK."

Results for SPSS's Frequencies procedure are shown in Figure 3.4. Similar output is obtained by using other software packages. This figure confirms that the data set includes 400 cases. Figure 3.4 also shows descriptive statistics for each of the three variables in the data set. Note that the mean for tenure status is .60. Because the coding was such that tenured faculty received a 0 and untenured faculty a 1, a mean of .60 indicates that the sample includes 40% (i.e., 160) tenured individuals. Figure 3.4 also shows that faculty received an annual salary increase ranging from a low of \$13.72 to a high of \$2,148.91. Overall performance scores ranged from a minimum of 1.00 to a maximum of 5.00.

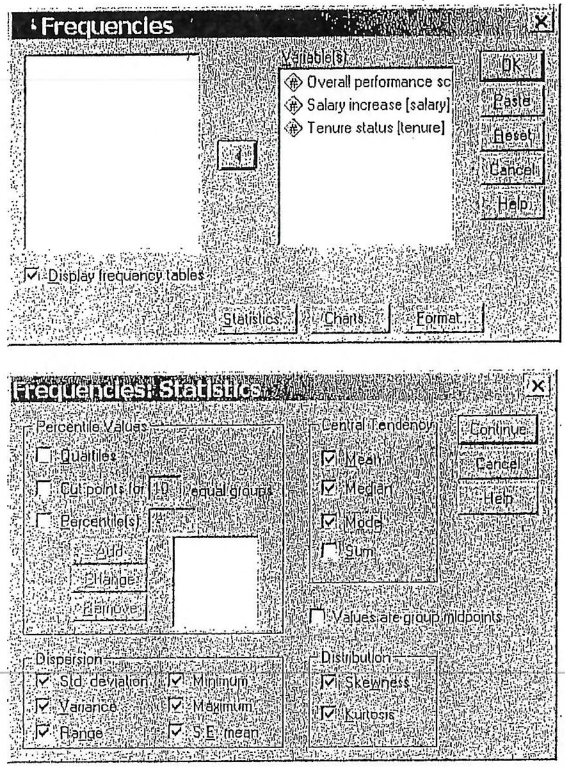


FIGURE 3.2. SPSS screens for the Frequencies procedure (top) and options for Frequencies: Statistics (bottom).

Next is a description of the two steps needed to conduct the MMR analysis and test whether tenure is a moderator of the performance score–salary increase relationship.

CONDUCTING AN MMR ANALYSIS USING COMPUTER PROGRAMS: TWO STEPS

Step 1: Computation of Product Term

Recall that Chapter 2 described that we need to form two regression equations, one including the first-order effects only and a second (i.e., MMR model) including the first-order effects as well as a product term

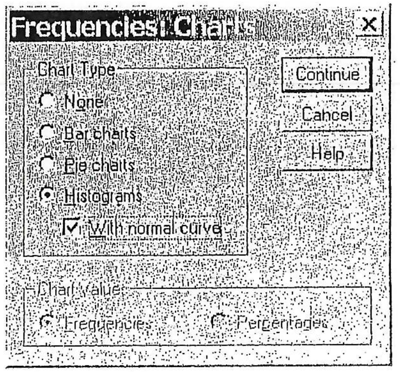
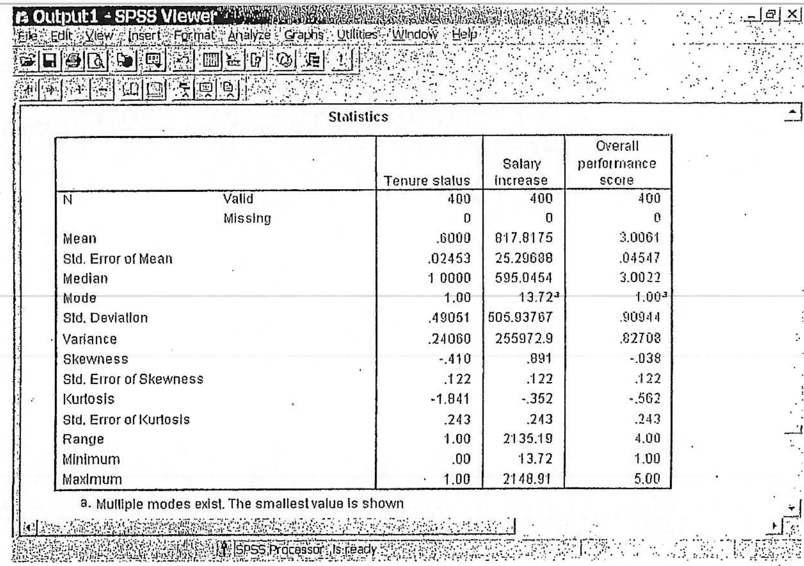


FIGURE 3.3. SPSS screen for Frequencies: Charts.

including the hypothesized moderator variable. In our research scenario, the product term is *Perf* × *Tenure*. Thus, we first need to create the product term. To do so, we create a new variable consisting of the product between *Perf* and *Tenure*. In SPSS, from the "Transform" pull-down menu, go to the "Compute" option. In the "Target Variable" window,



Statistics

		Tenure status	Salary increase	Overall performance score
N	Valid	400	400	400
	Missing	0	0	0
Mean		.6000	817.8175	3.0061
Std. Error of Mean		.02453	25.29688	.04547
Median		1.0000	595.0454	3.0022
Mode		1.00	13.72ᵃ	1.00ᵃ
Std. Deviation		.49051	505.93767	.90944
Variance		.24060	255972.9	.82708
Skewness		-.410	.891	-.038
Std. Error of Skewness		.122	.122	.122
Kurtosis		-1.841	-.352	-.562
Std. Error of Kurtosis		.243	.243	.243
Range		1.00	2135.19	4.00
Minimum		.00	13.72	1.00
Maximum		1.00	2148.91	5.00

a. Multiple modes exist. The smallest value is shown

FIGURE 3.4. SPSS output screen for the Frequencies procedure.

enter the name for the product term, for example, "perfxten." Then, in the "Numeric Expression" window, enter "perf * tenure." Figure 3.5 shows the Compute Variable screen once all this information has been entered. Now, click "OK."

The data screen should now include a fourth column with the newly created *perfxten* variable. Figure 3.6 shows the updated data screen.

Step 2: Computation of Regression Equations

As described in Chapter 2, the equations that need to be formed are the following (error terms are omitted for the sake of simplicity):

$$Salary = a + b_1 Perf + b_2 Tenure \tag{3.1}$$

$$Salary = a + b_1 Perf + b_2 Tenure + b_3 Perf \cdot Tenure \tag{3.2}$$

To compute these equations, we need to implement the regression procedure. In SPSS, from the "Analyze" pull-down menu, click on the "Regression" and "Linear" choices. The criterion Salary goes in the "Dependent" variable window, and the predictors *Perf* and *Tenure* go in the "Independent" variable window. Figure 3.7 (top) shows the corresponding SPSS screen.

Note that Figure 3.7 (top) shows that we entered variables *Perf* and *Tenure* in "Block 1 of 1." We now need to enter the product term into the equation to compute Equation 3.2. We do so by clicking on "Next"

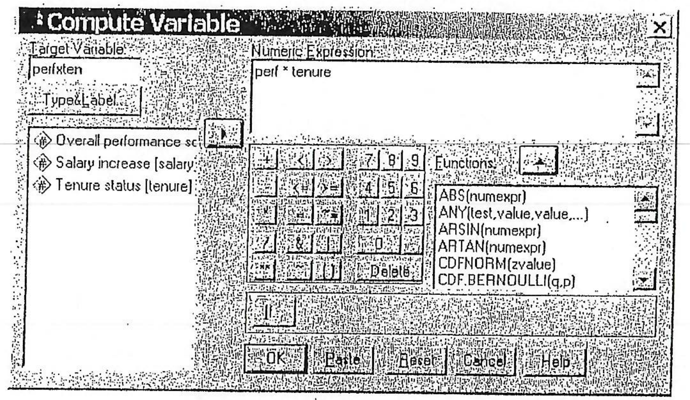


FIGURE 3.5. SPSS screen for creating the product term that carries information regarding the moderating effect of tenure.

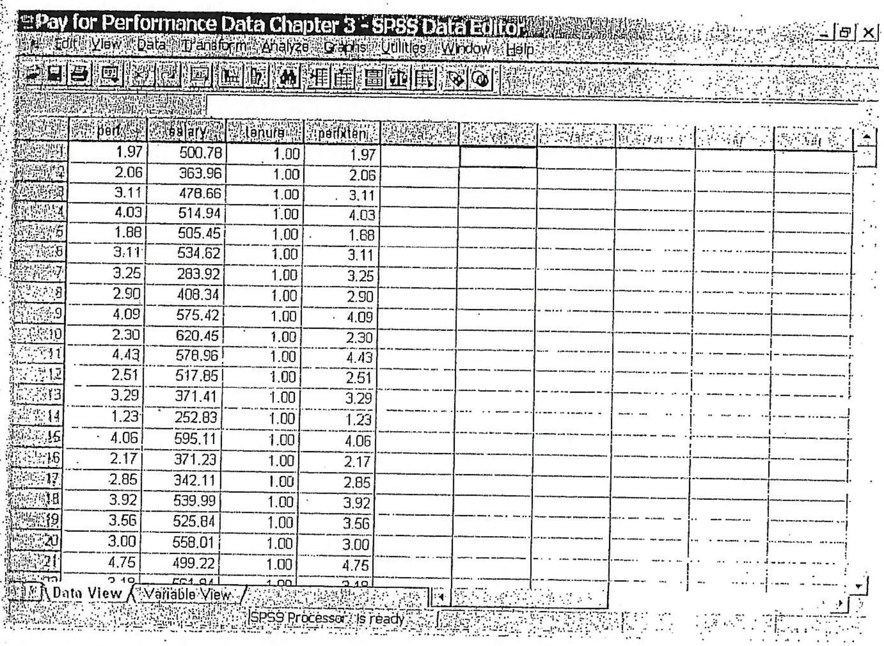


FIGURE 3.6. SPSS data screen for the Pay for Performance data set including the newly created *perfxten* variable (i.e., *Perf* · *Tenure*).

and placing the variable "perfxten" in the Independent variable window. This is shown in Figure 3.7 (bottom).

As noted in Chapter 2, the product term must be entered after its components are already in the equation. If the product term is entered before its components, it is likely to artificially inflate the size of the moderating effect (Stone, 1986, 1988; Tisak, 1994). It is also acceptable to enter all three variables (i.e., X, Z, and the product term) simultaneously in Block 1. However, entering all three predictors as part of Block 1 will not allow the computer program to generate the difference in R^2s between the model with the first-order effects only (i.e., Equation 3.1) and the model with the first-order effects and the product term (i.e., Equation 3.2). Therefore it is advisable to enter the first-order effects first and the product term second.

Finally, once the predictor and the moderator are entered in Block 1 and the product term is entered in Block 2, computer programs allow for several output options. In SPSS, for the "Statistics" options, we can click on estimates and confidence intervals for "Regression Coeffi-

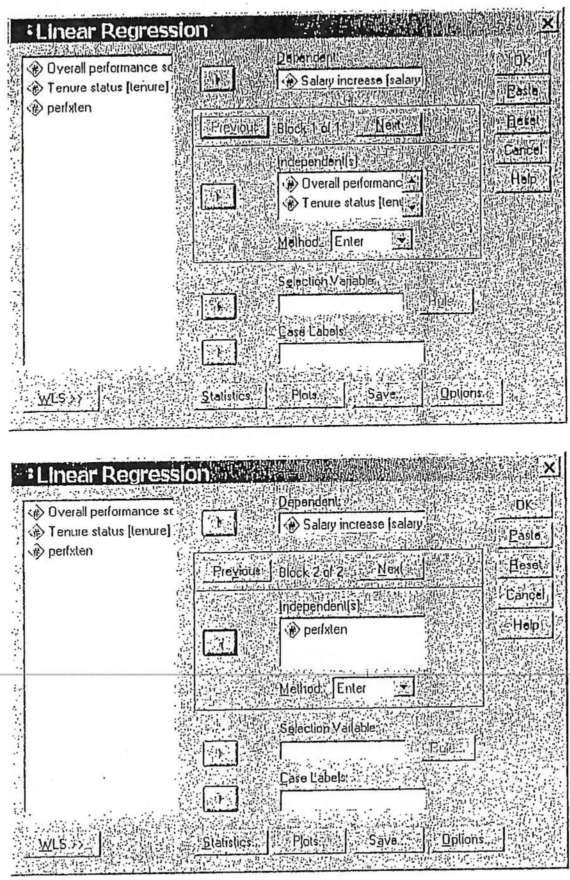


FIGURE 3.7. SPSS screen showing the computation of Equation 3.1 (top) and the addition of variable *perfxten* in Block 2 to compute Equation 3.2 (bottom).

cients," and click on "Model fit," "R squared change," and "Descriptives." Figure 3.8 shows this SPSS screen.

Regarding "Plots" options in Figure 3.7, it is useful to produce all standardized residuals plots to check for compliance with the ordinary least-squares assumptions described in Chapter 2 (e.g., homoscedasticity, normality of residuals; see Cohen et al., 2003, pp. 125–141). Then, click on "OK" to run the procedure.

OUTPUT INTERPRETATION

Interpretation of Model 1

All major computer packages include output similar to the information produced by SPSS and shown in Figure 3.9. Figure 3.9 (top) shows that for Model 1 (i.e., Equation 3.1) $R = .925$, $R^2 = .855$, and $F(2, 397) = 1168.67$, $p = .000$. This R^2 means that 85.5% of the variance in salary increase is explained by performance scores and tenure status. Specifically, the Coefficients table from the SPSS output reproduced in Figure 3.9 (bottom) shows that the resulting regression equation for Model 1 is the following:

$$\textit{Predicted Salary} = 669.08 + 222.76\,\textit{Perf} - 868.14\,\textit{Tenure} \quad (3.3)$$

Figure 3.9 (top) also indicates that the adjusted $R^2 = .854$. This statistic attempts to correct for capitalization on chance by applying a "correction" factor to R^2 based on the size of the sample and the number of predictors included in the regression model. The smaller the sample and the greater the number of predictors, the smaller the adjusted R^2 (St. John & Roth, 1999). In this illustration, the difference between R^2 and adjusted R^2 is very small because sample size is quite large (i.e., $N = 400$) and the regression equation includes two predictors only.

The coefficients for both performance and tenure in Model 1 are statistically significant at the $p < .001$ level. Equation 3.3 shows that for

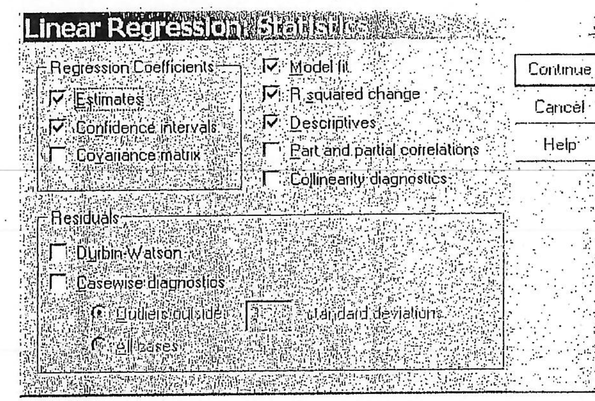


FIGURE 3.8. SPSS screen showing the Linear Regression: Statistics options.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.925[a]	.855	.854	193.26707	.855	1168.666	2	397	.000
2	.957[b]	.917	.916	146.72687	.062	292.790	1	396	.000

a. Predictors: (Constant), Tenure status, Overall performance score

b. Predictors: (Constant), Tenure status, Overall performance score, PERFXTEN

Coefficients[a]

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	t	Sig.	95% Confidence Interval for B: Lower Bound	Upper Bound
1	(Constant)	669.083	35.229		18.992	.000	599.824	738.342
	Overall performance score	222.756	10.641	.400	20.933	.000	201.836	243.676
	Tenure status	-868.138	19.729	-.842	-44.002	.000	-906.926	-829.351
2	(Constant)	183.246	39.006		4.698	.000	106.561	259.932
	Overall performance score	385.620	12.484	.693	30.888	.000	361.077	410.164
	Tenure status	-27.799	51.344	-.027	-.541	.589	-128.740	73.142
	PERFXTEN	-280.196	16.375	-.906	-17.111	.000	-312.389	-248.003

a. Dependent Variable: Salary increase

FIGURE 3.9. SPSS output screen showing Model 1 and Model 2 summary statistics (top) and regression coefficients (bottom).

a 1-point increase in performance score, salary is predicted to increase by $222.76, given that tenure is held constant (i.e., this is the average predicted increase in salary per 1-unit increase in performance across groups). Thus, an improvement in performance of 4 points is likely to lead to an increase of about $900 across groups. This may not be considered a "practically significant" effect for any one given year. However, small salary increases can be meaningful over a 30-year career span (also, a more detailed discussion regarding ways to assess practical significance is provided in Chapter 9). The regression coefficient associated with *Tenure* means that the difference in salary increase between the tenured and untenured groups is $868.14, given that performance score is held constant (i.e., this is the predicted difference in salary between a tenured versus untenured faculty member, assuming their performance scores are equal). Because dummy coding was used and the tenured group received a value of 0, the intercept in Equation 3.3 means that a tenured faculty with an average performance score across the entire sample (i.e., M_{Perf} = 3.0061; see Figure 3.4) is predicted to receive a salary increase of $669.08. Note that the way the data were coded has important implications regarding result interpretation. Specifically, had we used the 0 value for the untenured group and 1 for the tenured group, the intercept would represent the predicted salary increase for an untenured faculty member with an average performance score.

Model 1 does not include the product term and, thus, ignores a possible moderating effect of tenure status. In other words, this model shows that, holding tenure status constant, salary increases by an average of $222.76 when performance increases 1 point. However, so far we have no information regarding the potential moderating effect of tenure on the performance score–salary increase relationship. Could it be that the effect of performance score on salary increase is contingent on tenure status? The answer to this question is given by interpreting Model 2.

Interpretation of Model 2

Model 2 (i.e., Equation 3.2) shows results after the product term has entered the equation. As shown in Figure 3.9 (top), the addition of the product term resulted in an R^2 change of .062, $F(1, 396) = 292.79$, $p < .001$. This result supports the presence of a moderating effect. In other words, the moderating effect of tenure explains 6.2% of variance in salary increase above and beyond the variance explained by performance scores and tenure status.

The SPSS output shown in Figure 3.9 (bottom) also includes information regarding the regression coefficients after the product term is entered in the equation. The equation is the following:

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80\,\textit{Tenure} - 280.20\,\textit{Perf} \cdot \textit{Tenure} \quad (3.4)$$

This output screen also shows that, as described in Chapter 2, the statistical significance (i.e., p value) for the R^2 change from Model 1 to Model 2 based on the F statistic is identical to the statistical significance for the regression coefficient for the product term based on the t statistic (i.e., $p < .001$).

Once again, we base the interpretation of the regression coefficients on the fact that we coded the binary moderator using the dummy coding system. This coding scheme is recommended for situations involving binary moderators because of its simplicity and ease of interpretation of the results. The interpretation of the regression coefficient for the product term in Equation 3.4 is that there is a –$280.20 difference between the slope of salary increase on performance between the untenured (coded as 1) and the tenured group (coded as 0). In other words, the slope regressing salary on performance is less steep for untenured faculty members as compared to tenured faculty members.

Because the interpretation of the coefficient of the product term can be confusing, it is typically useful to create a graph displaying the performance–salary relationship for each of the groups. To do so, we first need to construct the regression equation for each of the two groups. Recall that tenured faculty were assigned a code of 0, whereas untenured faculty were assigned a code of 1. Therefore, reworking Equation 3.4 for the tenured group (i.e., *Tenure* = 0) yields the following:

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80\,\textit{Tenure} - 280.20\,\textit{Perf} \cdot \textit{Tenure} \quad (3.5)$$

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80(0) - 280.20(0)$$

Tenured Faculty: $\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf}$

Reworking Equation 3.4 for the untenured group (i.e., *Tenure* = 1) yields the following:

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80\,\textit{Tenure} - 280.20\,\textit{Perf} \cdot \textit{Tenure} \quad (3.6)$$

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80(1) - 280.20\,\textit{Perf}(1)$$

$$\textit{Predicted Salary} = 183.25 + 385.62\,\textit{Perf} - 27.80 - 280.20\,\textit{Perf}$$

$$\textit{Predicted Salary} = 155.45 + \textit{Perf}\,(385.62 - 280.20)$$

Untenured Faculty: $\textit{Predicted Salary} = 155.45 + 105.42\,\textit{Perf}$

Now, we can plot the performance–salary relationship for each group. To do so, it is recommended that we choose values of 1 standard deviation (*SD*) above and below the mean for *Performance* in Equations 3.5 and 3.6 (Cohen et al., 2003). Figure 3.4 shows that the mean score for *Performance* is 3.00, and the *SD* is .91. So, using the value of 3.91 (1 *SD* above the mean) and 2.09 (1 *SD* below the mean) in Equations 3.5 and 3.6 yields the graph shown in Figure 3.10.

Results based on Equation 3.4 led to the conclusion that there is a moderating effect of tenure. A perusal of Figure 3.10 showing the performance score–salary increase relationship for each of the groups separately gives us a better sense that the relationship is stronger (i.e., steeper slope) for the tenured faculty as compared to the untenured faculty group.

Additional Issues in Output Interpretation

Results indicate a moderating effect such that the relationship between performance and salary increase is stronger for tenured as compared to untenured faculty members. In some situations, it may be useful to interpret the coefficient associated with performance score, ignoring the moderating effect (i.e., Model 1). Because we are not taking into account the interaction effect, this coefficient can be considered an *average effect* of the relationship between performance and

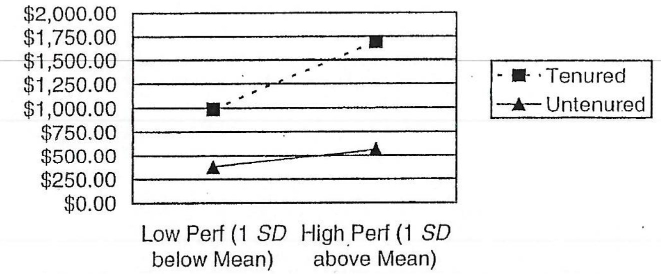


FIGURE 3.10. Slopes for *Salary* on *Performance* for tenured and untenured faculty based on Equations 3.5 and 3.6.

salary across levels of the moderator variable (i.e., tenure status) (Jaccard et al., 1990; Overall, Lee, & Hornick, 1981). The presence of the interaction implies that this average was computed from heterogeneous values (i.e., a larger coefficient for tenured than untenured faculty). Therefore, the interpretation of the interaction yields more detailed and precise information about the relationship between performance and salary for various groups (i.e., tenured vs. untenured faculty). On the other hand, the interpretation of the first-order effects provides less precise information.

Additional considerations are needed if we wish to interpret the intercept and the coefficient associated with the first-order effect of the predictor hypothesized to be the moderator in the presence of a nonzero moderating effect in Model 2 (Katrichis, 1993). Consider Equation 3.4, including the moderating effect of tenure. The interpretation of the regression coefficient for *Tenure* is that the estimated difference between the salary increase of an untenured faculty member and a tenured one, both with a performance score of 0, is –$27.80. But, a score of 0 does not even exist for *Performance* because this variable was measured on a scale ranging from 1 to 5! Thus, the regression coefficient for *Tenure* in Equation 3.4 is not really meaningful. Similarly, the intercept in the model including the product term is interpreted as the salary increase for a member of the tenured group (because tenure was coded as 0) with a performance score of 0. Again, this is not very meaningful because the value of 0 falls outside of the 1–5 scale used to measure performance. Note, however, that the coefficient for the moderator and the intercept are more readily interpretable when the predictor *X* scale includes a meaningful zero point.

One way to make the first-order coefficient for the moderator and the intercept more interpretable is to center the quantitative predictors (Aiken & West, 1991, Chapter 3). Centering achieves the goal of making zero a meaningful value and, consequently, the coefficients become easier to understand. Centered scores are obtained by simply subtracting the mean from each score, resulting in transformed scores with a mean = 0 (Tate, 1984). It should be noted, however, that in some situations the MMR models based on centered and uncentered predictors may be functionally equivalent (Kromrey & Foster-Johnson, 1998). Also, centering is just one of several possible ways of making the zero point more meaningful. An alternative procedure includes transforming the zero point into the scale midpoint or a neutral midpoint; for example, $Performance_{mid} = Performance - 3$ (because in this case 3 is the midpoint on the scale used to measure *Performance*). Another alternative is to transform the zero point into the median: $Performance_{med} = Performance - MDN_{Performance}$. In the latter case, the coefficient associated with

Tenure is interpreted as the effect of *Tenure* on *Performance* at the median score for *Performance*.

A simple way to perform the centering procedure using computer packages is to create a new "centered performance" variable. In SPSS, this is achieved by using the "Compute" command. Using our data set, the centered variable is obtained as follows:

$$\textit{Centered Performance } [cperf] = Perf - 3.0061 \qquad (3.7)$$

As a way to make sure the procedure has been conducted in the appropriate way, it is a good idea to run the Frequencies procedure and obtain descriptive statistics for the centered variables. If centering was done correctly, the mean score for *cperf* should be 0. Then the centered variables and the product between the centered variables are used to rerun the MMR analyses. The resulting MMR model is the following:

$$\textit{Predicted Salary} = 1342.46 + 385.62\ cperf - 870.10\ tenure - 280.20\ cperf \cdot tenure \qquad (3.8)$$

Based on Equation 3.8, the intercept means that the estimated salary increase for a tenured faculty member with an average performance score (i.e., 3.00) is $1,342.46, taking into account the performance–salary increase relationship for the tenured group. The regression coefficient associated with tenure means that the difference in salary increase between an untenured and a tenured faculty member with average performance scores is –$870.10. The regression coefficient associated with centered performance means that the salary increase for a tenured faculty member with an average performance score is $385.62. This is identical to the slope that would be obtained if we regressed salary increase on performance for the group of tenured faculty only (i.e., for tenured faculty *Predicted Salary* = 183.25 + 385.62 *Perf*). Again, the interpretation of these coefficients is guided by the fact that dummy coding was used.

A comparison of Equation 3.4, using uncentered performance scores, with Equation 3.8, using centered performance scores, reveals that there are important changes in the intercept as well as the coefficient associated with the effect of *Tenure*. However, the coefficients associated with the quantitative predictor and the product term have remained unchanged (Lautenschlager & Mendoza, 1986). Also, the R^2s associated with Model 1 and Model 2 are identical to the situation including uncentered predictors. This is because simple additive transformations on predictor variable scores such as centering do not change the statistical test of the product term (Arnold & Evans, 1979; Cohen, 1978; Friedrich, 1982). Similarly, the R^2s associated with Model 1 and Model 2 are not affected by the use of different coding schemes.

Note that there is no need to center the quantitative criterion (i.e., salary increase, in this example). If the criterion is in its original uncentered scale, then predicted scores are also expressed in the same units as the original scale. Thus, the interpretation of predicted criterion scores is more straightforward when left in their original metric.

Transformations involving adding or subtracting a constant such as centering also do not affect regression coefficients in Model 1 (i.e., Equation 3.3) (but they may affect the intercept). As noted earlier, they do not affect the coefficient for the product term in Model 2 (i.e., Equation 3.4) (Irwin & McClelland, 2001). However, as seen earlier, these types of transformations have dramatic effects on the coefficients of the lower-order terms because coefficients resulting from uncentered variables are referenced to a zero point on the other predictor, whereas coefficients resulting from centered variables are referenced to the *average* value on the other predictor. Nevertheless, the interpretation of the moderating effect is identical in equations based on centered and uncentered predictors. Consequently, as long as the scales have interval-level properties and arbitrary zero points (i.e., Likert-type scales) (Aguinis, Henle, & Ostroff, 2001), it is still appropriate to test moderating effects.

It should be emphasized that interpretation of the intercept and the regression coefficient associated with the moderator is likely to change if a different coding scheme is used (Schoorman, Bobko, & Rentsch, 1991). Because each coding scheme represents a different way to partition the total variance associated with the interaction, the regression coefficients associated with the first-order effects and the intercept answer a different question for each coding scheme and, therefore, their value and statistical significance level also vary (Schoorman et al., 1991). However, the value and significance level for the coefficient for the product term as well as the model's R^2 are not affected by the use of a different coding type. A more detailed discussion of the effects of various coding strategies on the interpretation of the results is provided in Chapter 8.

Finally, in interpreting the computer output for Model 2, MMR researchers are better served by examining the unstandardized regression coefficients (i.e., *bs*) as opposed to the standardized coefficients (i.e., betas) (Friedrich, 1982; Jaccard et al., 1990). The interpretation of the "standardized" solution provided by computer packages is fraught with great difficulties because even if the predictors are standardized, their product is not necessarily standardized. This is because instead of first standardizing X and Z (i.e., Z_X, Z_Z) and then obtaining their product (i.e., $Z_X \cdot Z_Z$), computer programs first generate the product between X and Z and then standardize the resulting scores (i.e., $Z_{X \cdot Z}$). Because

$Z_X \cdot Z_Z$ and $Z_{X \cdot Z}$ may differ, the "standardized" solution produced by computer outputs may be misleading. If an MMR user is interested in obtaining regression coefficients that are interpretable in a standardized metric, this is possible by first converting each of the predictors and the criterion into standard scores (i.e., [score – mean]/*SD*), and then creating the product term based on the standard scores for X and Z (i.e., $Z_X \cdot Z_Z$). Then, because all variables have been converted to standard scores, the *bs* shown in the computer output can be interpreted using a standard score metric (Friedrich, 1982; Jaccard et al., 1990).

Figure 3.11 shows the SPSS output screen corresponding to the MMR analysis including standardized scores for *Salary*, *Performance*, and *Tenure*. For Model 2, the interpretation of the unstandardized coefficient associated with *Standardized Tenure* is that the difference in standardized salary increase between an untenured and a tenured faculty member with average performance scores is .844. The interpretation of the unstandardized coefficient for *Standardized Performance* is that the mean standardized salary increase for a tenured faculty member with an average performance score is .39. Finally, the unstandardized coefficient associated with the product terms means that the standardized difference for the slope of salary increase on performance between the untenured (coded as 1) and the tenured group (coded as 0) is –.247. In other words, the slope is steeper for the tenured group by .247 salary-increase standard deviation units.

CONCLUSIONS

- This chapter provided a step-by-step description of how to conduct an MMR analysis including a binary moderator variable using computer programs, with a special emphasis on SPSS. The procedure involves creating a new variable that consists of the product term between the predictor and the moderator variables, and implementing a hierarchical regression procedure. This chapter described how to interpret the computer output, including answering the key issue of whether the Y on X slope differs across the moderator-based subgroups.

- The chapter used dummy coding for the binary moderator. In this coding scheme, members of one of the groups are arbitrarily assigned a 0 and members of the other group are assigned a 1. This coding scheme is recommended for situations involving binary moderators because of its simplicity and ease of interpretation of the results. The choice of a coding scheme affects the interpretation of the intercept and the regression coefficient associated with the first-

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.860E-05	.019		.001	.999	-.038	.038
	ZPERF	.400	.019	.400	20.933	.000	.363	.438
	ZTENURE	-.842	.019	-.842	-44.002	.000	-.879	-.804
2	(Constant)	5.110E-03	.015		.352	.725	-.023	.034
	ZPERF	.391	.015	.391	26.904	.000	.362	.420
	ZTENURE	-.844	.015	-.844	-58.088	.000	-.872	-.815
	ZPRODUCT	-.247	.014	-.249	-17.111	.000	-.275	-.219

a. Dependent Variable: ZSALARY

FIGURE 3.11. SPSS output screen showing Model 1 and Model 2 regression coefficients based on standardized scores.

order effect of the moderator in the MMR model. Coding schemes other than dummy coding represent a different way of partitioning the total variance associated with the interaction. Therefore, the regression coefficients associated with the first-order effects and the intercept answer a different question for each coding scheme and, consequently, their value and statistical significance level also vary. On the other hand, the value and significance level for the coefficient for the product term as well as the model's R^2 are not affected by a change in the coding scheme.

- Centering of the predictor X is an additional issue that should be taken into account if a researcher wishes to interpret the first-order coefficients in the presence of a nonzero interaction. The coefficient for the first-order effect of the moderator variable is referenced to a *zero* point on the other predictor, and this "zero point" may not be meaningful given the scale used to measure the predictor X. On the other hand, the coefficient resulting from centering the X variable is referenced to the *average* value on X. Thus, implementing an additive transformation on X such as centering is likely to result in a more meaningful coefficient for the first-order effect of the moderator. Note, however, that the size, statistical significance, and interpretation of the coefficient associated with the product term are identical in equations based on centered and uncentered predictors.

- Going back to the research questions posed at the beginning of this chapter, the finding is that there is an overall positive relationship between performance scores and salary increase. Regarding the substantive moderating-effect hypothesis, tenure status is a moderator such that the relationship is stronger for tenured than for untenured faculty. In practical terms, untenured faculty do not receive as high a pay increase as compared to tenured faculty given the same performance score. This is a critical piece of information in implementing the new performance management system. It is likely that untenured faculty will perceive the system to be unfair. Therefore, the system may have an effect exactly opposite to what was intended: The new performance management system may serve as a factor that decreases the motivation and satisfaction of untenured faculty.

- The next chapter addresses a critical statistical assumption that, if violated, may bias conclusions regarding the presence of a moderating effect. Specifically, Chapter 4 addresses a fundamental yet often ignored assumption of MMR: Homogeneity of error variance.