Biometrika (1989), **76**, 1, pp. 81-92
Printed in Great Britain

Moderating effects of subgroups in linear models

BY NANNY WERMUTH
Psychologisches Institut, Johannes Gutenberg-Universität, 6500 Mainz, Federal Republic of Germany

SUMMARY

Possibilities for moderating effects of a subgrouping variable on strength or direction of an association have been much discussed by social scientists but have not been given satisfactory statistical formulations. The results concern directed measures of associations in linear models containing just three variables.

Some key words: Analysis of covariance; Analysis of variance; CG-distribution; Conditional independence; Graphical chain model; Parallel regressions; Yule-Simpson paradox.

1. INTRODUCTION

Linear models are commonly used as a framework to estimate and test how a continuous response variable depends on potential influencing variables. This paper is concerned with the situation in which two influences are random variables as well, which may be discrete or continuous. The continuous random variables are to stand for quantitative properties of observational units, while the discrete variables either capture qualitative properties or they represent subgroups of the population to which the observational units belong.

To think of all variables as random variables is usually appropriate for data obtained from observational studies, where one cannot control which levels or values of the influencing variables are to be observed. The linear model permits an analysis conditional on fixed levels and values of the influences. But frequently, the association structure among the influences is, in social science applications, of interest itself and, as it will turn out, it is important for a correct interpretation of how the response depends on the influences.

Awareness of such problems in interpretation is widespread among social scientists as is documented by extensive discussions of moderating effects; see, for example, Saunders (1956), Zedeck (1971) and Baron & Kenny (1986). There, a variable is called a moderator if its presence changes the strength or direction of an association. However, methods recommended to social scientists for identifying moderating effects given in the above literature or by Cohen & Cohen (1983, pp. 310-4) have been shown to be seriously deficient (Wermuth, 1988). Thus, there is need for clarification. We treat the case of a potential discrete moderator variable and call its categories or levels the subgroups in the linear model.

A moderating effect is closely tied to the notion of consistent results on an association (Wermuth, 1987). In the present context results are said to be weakly consistent if the associations within subgroups coincide in direction and strongly consistent if the associations within subgroups coincide in direction and strength. This notion depends typically on the chosen measure of association. In linear models we look at directed measures of associations. These are regression coefficients if the influence is continuous, and they are effect parameters if the influence is discrete.

A discrete variable has a moderating effect on a measure of association either if the results are not strongly consistent within subgroups, or if the results are strongly consistent but do not coincide with the overall results obtained after pooling over the subgroups. Expressed in statistical terminology this says that a moderating effect is lacking if equal partial associations coincide with a corresponding marginal association or, to put it differently, if homogeneous associations at fixed levels of a third variable are collapsible over this variable.

Extreme situations for noncollapsible associations, where, for instance, positive partial associations coexist with a corresponding negative marginal association are known as the Yule–Simpson paradox (Yule, 1900; Simpson, 1951) for only discrete variables, and have been described for effect parameters in linear models, as well (Snedecor & Cochran, 1967, p. 472). Necessary and sufficient conditions for the lack of a moderating effect have been given for different types of measure of association in the case of only continuous variables with exclusively linear relations (Wermuth, 1989), and in the case of only discrete variables for a symmetric measure of association (Whittemore, 1978), as well as for a directed measure of association (Wermuth, 1987).

In §§ 2 and 3, conditions are derived for the lack of a moderating effect of a discrete variable on directed measures of associations in linear models with a continuous response. These are conditions on the association structure of the variables under study. Thus, this paper is concerned with properties of the target population, not with properties of the design or execution of a study, nor with properties of estimates. In § 4 the results are applied to two sets of data from psychological research.

2. Moderating effects on a homogeneous regression coefficient

Here we look at the linear dependence of a continuous response X, a continuous influence Y and a potential discrete moderator A having levels $i=1,\ldots,I$. We assume a homogeneous linear dependence of X on Y in terms of regression coefficients and constant residual variances (2·1), a possible dependence of the means of Y, but not of the variances of Y on A (2·2) and for A a probability function, which is arbitrary except for positive probabilities at all levels (2·3);

$$E(X \mid Y=y, A=i)=\alpha_{xy}(i)+\beta_{xy}y, \quad \operatorname{var}(X \mid Y=y, A=i)=\tau_{xx.y}, \tag{2·1}$$

$$E(Y \mid A=i)=\mu_y(i), \quad \operatorname{var}(Y \mid A=i)=\sigma_{yy}, \tag{2·2}$$

$$\operatorname{pr}(A=i)=\pi_i>0. \tag{2·3}$$

Equivalently to (2·1) and (2·2), we could have specified that the conditional mean and variance of the random vector $Z=[X \;\; Y]$ given A is

$$E(Z \mid A=i)=[\mu_x(i) \;\; \mu_y(i)], \quad \operatorname{var}(Z \mid A=i)=\begin{bmatrix}\sigma_{xx} & \sigma_{xy}\\ \sigma_{yx} & \sigma_{yy}\end{bmatrix}, \tag{2·4}$$

with

$$\mu_x(i)=\alpha_{xy}(i)+\beta_{xy}\mu_y(i), \quad \sigma_{xy}=\beta_{xy}\sigma_{yy}, \quad \sigma_{xx}=\tau_{xx.y}+\beta_{xy}^2\sigma_{yy}.$$

By adding the assumption of a normal distribution for X given Y and A equations (2·1) define an analysis of covariance model, which is also known as a parallel regression model. If we assume, in addition, a normal distribution for Y given A, then (2·1) to (2·3) give one parameterization of a saturated homogeneous CG-distribution for X, Y and A; CG-distributions have been studied by Lauritzen & Wermuth (1989). They show, for instance, that the density g_{AXY} of a joint CG-distribution for X, Y and A can be factorized into densities of distributions, which are all of the CG-type as

$$g_{AXY} = g_{XY|A}g_A = g_{X|YA}g_{Y|A}g_A. \tag{2·5}$$

One speaks of a CG-distribution, since the continuous variables have a conditional Gaussian distribution given the discrete variables, and the distribution of the discrete variables is taken to be arbitrary except for having positive probabilities everywhere. It is called a homogeneous CG-distribution if only the means but not the covariance matrix of the continuous variables may depend on the levels of A; compare (2·4). A CG-distribution is saturated if constraints on its interaction terms just ensure uniqueness of the interactions, but do not introduce restrictions on the specified distribution; see Appendix 2.

Some of the conditions derived below for the lack of a moderating effect of A on β_{xy} may be restated in terms of conditional independencies if X, Y, A have a homogeneous CG-distribution. In this case we know the following.

Facts 1.

(a) X is conditionally independent of Y given A, that is $X \perp\!\!\!\perp Y \mid A$, if and only if $\beta_{xy} = 0$;

(b) X is conditionally independent of A given Y, that is $X \perp\!\!\!\perp A \mid Y$, if and only if $\alpha_{xy}(i) = \alpha_{xy}$;

(c) Y is independent of A, that is $Y \perp\!\!\!\perp A$, if and only if $\mu_y(i) = \mu_y$.

The homogeneous regression coefficient β_{xy} in (2·1) is collapsible over A if it coincides with the marginal regression coefficient $\tilde{\beta}_{xy}$ defined by

$$E(X \mid Y = y) = \tilde{\alpha}_{xy} + \tilde{\beta}_{xy} y,$$

and in terms of the marginal covariance $\tilde{\sigma}_{xy}$ and the marginal variance $\tilde{\sigma}_{yy}$ as $\tilde{\beta}_{xy} = \tilde{\sigma}_{xy}/\tilde{\sigma}_{yy}$. By computing the marginal moments from the conditional ones in (2·4) one obtains, as shown in Appendix 1, the quantification of the moderating effect of A on β_{xy} as

$$\tilde{\beta}_{xy} - \beta_{xy} = [\textstyle\sum \pi_i\{\alpha_{xy}(i) - \tilde{\alpha}_{xy}\}\{\mu_y(i) - \tilde{\mu}_y\}]/\tilde{\sigma}_{yy}. \tag{2·6}$$

This permits us to state the following as an immediate consequence.

PROPOSITION 1. *In a parallel regression model for X on Y and A given by* (2·1) *with random influences characterized by* (2·2) *and* (2·3):

(i) *the homogeneous regression coefficient is collapsible over A, $\beta_{xy} = \tilde{\beta}_{xy}$, if and only if*

$$\mathrm{cov}_A\{\alpha_{xy}(i), \mu_y(i)\} = 0;$$

(ii) *variable A has no moderating effect on the linear dependence of X on Y if the parallel regressions are also coincident, $\alpha_{xy}(i) = \alpha_{xy}$, or if the means of the influence Y coincide in the subgroups, $\mu_y(i) = \mu_y$.*

Remark 1. The sufficient condition in Proposition 1(ii) is also necessary if the discrete variable A is dichotomous, $I = 2$.

Remark 2. If X, Y and A follow a homogeneous CG-distribution then the sufficient condition in Proposition 1(ii) can be restated in terms of independencies as $X \perp\!\!\!\perp A \mid Y$ or $Y \perp\!\!\!\perp A$. This is a consequence of Facts 1. If only one of these independencies hold, then the marginal joint distribution of X and Y is not Gaussian (Yakowitz & Spagins, 1968). This just illustrates that parametric collapsibility does not imply collapsibility in distribution.

Remark 3. If X, Y and A follow a nonhomogeneous CG-distribution, but the regression lines of X on Y are, nevertheless, parallel in all subgroups, $\beta_{xy}(i) = \beta_{xy}$, then A still has no moderating effect on the linear dependence of X on Y if $X \perp\!\!\!\perp A \mid Y$ or $Y \perp\!\!\!\perp A$. The reasons are that:

(i) for the nonhomogeneous CG-distribution we know $X \perp\!\!\!\perp A \mid Y$ if and only if $\tau_{xx.y}(i) = \tau_{xx.y}$ and $\beta_{xy}(i) = \beta_{xy}$ and $\alpha_{xy}(i) = \alpha_{xy}$, and $Y \perp\!\!\!\perp A$ if and only if $\sigma_{yy}(i) = \sigma_{yy}$ and $\mu_y(i) = \mu_y$;

(ii) given $Y \perp\!\!\!\perp A$ or $X \perp\!\!\!\perp A \mid Y$ we get $\text{cov}_A\{\alpha_{xy}(i),\, \mu_y(i)\} = 0$;

(iii) this is sufficient for the collapsibility of β_{xy} over A, since (2·6) shows the relation of β_{xy} to $\tilde{\beta}_{xy}$ even for a nonhomogeneous CG-distribution.

A stronger result can be stated if more is known about the association structure of X, Y and A: if $Y \perp\!\!\!\perp A \mid X$. This means that the potential moderator is not needed to predict values of the influence once knowledge about the response is available. Experience shows that quite a few sets of observable variables satisfy this condition even when the moderator and the influence are both important in predicting the response. The set of data in § 4·2 is one example.

PROPOSITION 2. *In a parallel regression model for Y on X and A derived from a homogeneous* CG*-distribution satisfying $Y \perp\!\!\!\perp A \mid X$, the following statements are equivalent*:

(i) *the discrete variable A has no moderating effect on the linear dependence of X on Y*;

(ii) *the response X is conditionally independent of the potential moderator A, that is $X \perp\!\!\!\perp A \mid Y$, or of the influence Y, that is $X \perp\!\!\!\perp Y \mid A$*;

(iii) *the influence Y is independent of the potential moderator A, that is $Y \perp\!\!\!\perp A$.*

Proof. The independence $Y \perp\!\!\!\perp A \mid X$ implies coincident regression lines of Y on X in all subgroups and, in particular, collapsible intercepts $\alpha_{yx}(i) = \tilde{\alpha}_{yx}$. The relation of these intercepts to means is $\alpha_{yx}(i) = \mu_y(i) - \beta_{yx}\mu_x(i)$. Together with equal intercepts it leads to

$$\mu_y(i) - \tilde{\mu}_y = \beta_{yx}\{\mu_x(i) - \tilde{\mu}_x\}. \tag{2·7}$$

The relation of $\alpha_{xy}(i)$ to means, (2·4), implies in general

$$\alpha_{xy}(i) - \tilde{\alpha}_{xy} = \{\mu_x(i) - \tilde{\mu}_x\} - \beta_{xy}\{\mu_y(i) - \tilde{\mu}_y\},$$

which turns with (2·7) into

$$\alpha_{xy}(i) - \tilde{\alpha}_{xy} = (1 - \rho_{xy}^2)\{\mu_x(i) - \tilde{\mu}_x\}, \tag{2·8}$$

where $0 \leq \rho_{xy}^2 = \beta_{yx}\beta_{xy} < 1$. Equations (2·7) and (2·8) permit us to write

$$\text{cov}_A\{\alpha_{xy}(i),\, \mu_y(i)\} = \beta_{yx}(1 - \rho_{xy}^2)\,\text{var}_A\{\mu_x(i)\}.$$

It follows from Proposition 1(i), (2·8) and Facts 1 that (i) is equivalent to (ii), and the equivalence of (ii) and (iii) is an immediate consequence of (2·7) and Facts 1. □

Since this proof does not depend on the distributional assumptions, Proposition 2 could have been stated in terms of conditions on parameters in (2·1)–(2·3), instead.

However, the formulation in terms of independencies points at analogies to situations in which the type of the involved variables is changed, as in § 3.

Remark 1. It is a direct consequence of Proposition 2 that for association structures with $Y \perp\!\!\!\perp A \mid X$, but not $X \perp\!\!\!\perp Y \mid A$, the conditions $X \perp\!\!\!\perp A \mid Y$ and $Y \perp\!\!\!\perp A$ are equivalent.

Remark 2. The results show that the conditional independence for pair (Y, A), which is typically of little interest in a conditional analysis of X given Y and A, becomes important for parametric collapsibility and moderating effects.

Remark 3. The assumptions for Proposition 2 may be tested in a stepwise manner within the framework of CG-distributions, as illustrated with the first example in § 4.

3. Moderating effects on a main effect parameter

Here we look at the dependence of a continuous response X on a discrete influence B having levels $j = 1, \ldots, J$ and a potential discrete moderator A having levels $i = 1, \ldots, I$. We assume a homogeneous dependence of the mean of X on B in terms of effect parameters and constant residual variances (3·1). For A and B a joint probability function is taken to be arbitrary except for positive probabilities at all level combinations of A and B (3·2)

$$E(X \mid A = i, B = j) = \mu_{ij} = \gamma_0 + \gamma_i^A + \gamma_j^B, \quad \text{var}(X \mid A = i, B = j) = \tau_{xx}, \tag{3·1}$$

$$\text{pr}(A = i, B = j) = \pi_{ij} > 0. \tag{3·2}$$

The terms in the linear expansion of μ_{ij} are called effect parameters after suitable constraints, like symmetric ones, have been imposed to ensure uniqueness, $\Sigma_i \gamma_i^A = \Sigma_j \gamma_j^B = 0$.

The main effects γ_j^B are the directed measures of partial associations for variable pair (X, B). By adding the assumption of a normal distribution for X given A and B, equations (3·1) define an analysis of variance model with only main effects γ_i^A and γ_j^B. By assuming, in addition, a joint distribution for A and B, equations (3·1) and (3·2) give one parameterization of a nonsaturated homogeneous CG-distribution for X, B and A with density

$$g_{XAB} = g_{X \mid AB}\, g_{AB},$$

having no linear three-factor interaction. This interaction, η_{ij}^{ABX}, discussed in Appendix 2, is zero, since it is a positive multiple of the two-factor interaction, γ_{ij}^{AB}, in the corresponding analysis of variance model and the latter is assumed to be zero with (3·1).

One consequence of having only main effects γ_i^A and γ_j^B in the conditional density $g_{X \mid AB}$ is a simple link between the marginal association of pair (A, B) in g_{AB} and the partial association of pair (A, B) in g_{XAB} as measured by the interaction parameters in a CG-distribution. After denoting this marginal and partial interaction by $\tilde{\lambda}_{ij}^{AB}$ and λ_{ij}^{AB}, respectively, one obtains, as is shown in Appendix 2,

$$\tilde{\lambda}_{ij}^{AB} = \lambda_{ij}^{AB} + \gamma_i^A \gamma_j^B / \tau_{xx}. \tag{3·3}$$

Some of the conditions derived below for the lack of a moderating effect of A on γ_j^B may be restated in terms of conditional independencies if X, B and A have a homogeneous CG-distribution and the interactive effect of A and B on the conditional mean of X given A and B is lacking. In this case we have the following.

Facts 2.
(a) X is conditionally independent of A given B, that is $X \perp\!\!\!\perp A \mid B$, if and only if $\gamma_i^A = 0$;
(b) A is conditionally independent of B given X, that is $A \perp\!\!\!\perp B \mid X$, if and only if $\lambda_{ij}^{AB} = 0$;
(c) A is independent of B, that is $A \perp\!\!\!\perp B$, if and only if $\tilde{\lambda}_{ij}^{AB} = 0$.

The main effect γ_j^B in (3·1) is collapsible over A if it coincides with the marginal effect $\tilde{\gamma}_j^B$ defined by

$$E(X \mid B = j) = \tilde{\mu}_j = \tilde{\gamma}_0 + \tilde{\gamma}_j^B, \quad \Sigma_j \tilde{\gamma}_j^B = 0. \tag{3·4}$$

By computing the marginal moments $\tilde{\mu}_j$ from the conditional ones in (3·1) one obtains, as shown in Appendix 1, the following expression for the moderating effect of A on γ_j^B:

$$\tilde{\gamma}_j^B - \gamma_j^B = \{\textstyle\sum_i \pi_{i|j} \gamma_i^A - \sum_j (\sum_i \pi_{i|j} \gamma_i^A)/J\}. \tag{3·5}$$

This permits us to state as immediate consequence the following.

Proposition 3. *In an analysis of variance model for X on A and B given by* (3·1) *with random influences characterized by* (3·2):
(i) *the main effect of B on X is collapsible over A*, $\gamma_j^B = \tilde{\gamma}_j^B$, *if and only if*

$$E_{A|B}(\gamma_i^A) = \textstyle\sum_j \{E_{A|B}(\gamma_i^A)\}/J;$$

(ii) *variable A has no moderating effect on the dependence of X on B if there is no main effect of the potential moderator A, that is* $\gamma_i^A = 0$, *or if the conditional probabilities for A given B do not depend on the levels of B, that is* $\pi_{i|j} = \pi_i$.

Remark 1. The sufficient condition in Proposition 3(ii) is also necessary if variables A and B are both dichotomous, $I = 2$ and $J = 2$.

Remark 2. If X, A and B follow a homogeneous CG-distribution, then the sufficient condition in Proposition 3(ii) can be restated in terms of independencies as $X \perp\!\!\!\perp A \mid B$ or $B \perp\!\!\!\perp A$. This is a consequence of Facts 2.

Remark 3. If X, A and B follow a nonhomogeneous CG-distribution, but there is, nevertheless, no interaction effect, $\gamma_{ij}^{AB} = 0$, in the analysis of variance model, then A has still no main effect on the dependence of the mean of X on B if $X \perp\!\!\!\perp A \mid B$ or $B \perp\!\!\!\perp A$. The reasons are that:
(a) for the nonhomogeneous CG-distribution we know that $X \perp\!\!\!\perp A \mid B$ if and only if $\tau_{xx}(i, j) = \tau_{xx}(j)$ and $\gamma_i^A = 0$, and that $A \perp\!\!\!\perp B$ if and only if $\pi_{i|j} = \pi_i$;
(b) given $A \perp\!\!\!\perp B$ or $X \perp\!\!\!\perp A \mid B$ we get $E_{A|B}\gamma_i^A = \Sigma_j \{E_{A|B}\gamma_i^A\}/J$, and
(c) this is sufficient for the collapsibility of γ_j^B over A, since (3·5) shows the relation of γ_j^B to $\tilde{\gamma}_j^B$ even for a nonhomogeneous CG-distribution.

A result similar to Proposition 2 can be stated if more is known about the association structure of X, B and A, if $B \perp\!\!\!\perp A \mid X$.

Proposition 4. *In an analysis of variance model for X and dichotomous A and B derived from a homogeneous* CG-*distribution satisfying* $B \perp\!\!\!\perp A \mid X$ *the following statements are equivalent*:
(i) *the main effect of influence B on the response X is not moderated by the discrete variable A*;

(ii) *the response X is conditionally independent of the potential moderator A, that is* $X \perp\!\!\!\perp A \mid B$, *or of the influence B, that is* $X \perp\!\!\!\perp B \mid A$;
(iii) *the influence B is independent of the potential moderator A, that is* $B \perp\!\!\!\perp A$.

Proof. If X, B and A have a homogeneous CG-distribution satisfying $B \perp\!\!\!\perp A \mid X$, then there is no interaction effect of A and B on the conditional mean of X given A and B. The equivalence of (ii) and (iii) follows then with (3·3) and Facts 2. The equivalence of (i) and $X \perp\!\!\!\perp A \mid B$ or $B \perp\!\!\!\perp A$ results from Proposition 3 and Remarks 1 and 2. By using the equivalence of (ii) to (iii), $X \perp\!\!\!\perp A \mid B$ or $B \perp\!\!\!\perp A$ is seen to imply (ii). Finally, (ii) implies (i) in the case of $X \perp\!\!\!\perp A \mid B$ with Proposition 3 and in the case of $X \perp\!\!\!\perp B \mid A$, since the independence of pair (X, B) is then not moderated by the presence of A. To put it differently, $X \perp\!\!\!\perp B \mid A$ and $B \perp\!\!\!\perp A \mid X$ imply $B \perp\!\!\!\perp (A, X)$ and, in particular, $X \perp\!\!\!\perp B$. □

Remark 1. It is a direct consequence of Proposition 4 that for association structures with $B \perp\!\!\!\perp A \mid X$, but not $X \perp\!\!\!\perp B \mid Y$, the conditions $X \perp\!\!\!\perp A \mid B$ and $B \perp\!\!\!\perp A$ are equivalent.

Remark 2. Proposition 4 implies that the common practice of using samples of equal size in two-way analysis of variance designs can be misleading, since taking equal numbers of observations for all level combinations of variables A and B forces these variables to behave like independent influences. More precisely, if the dichotomous influences and the response follow a homogeneous CG-distribution with $A \perp\!\!\!\perp B \mid X$ and strong associations of both influences to X, then:
(a) $\lambda_{ij}^{AB} = \eta_{ij}^{ABX} = 0$, while the interactions η_i^{AX} and η_j^{AY} are far from zero;
(b) there are high main effects of both variables in the corresponding conditional distribution of X given A and B, described with (3·1), since the main effects are simple multiples of the interactions η_i^{AX} and η_j^{AY}, for which see Appendix 2;
(c) the influences are marginally associated, as can be seen from (3·3) with $\lambda_{ij}^{AB} = 0$;
(d) each of the two main effects is moderated by the presence of the other variable.
It follows by Remark 1 that the important feature (b) of the target population cannot be detected in an analysis of variance based on a balanced sampling scheme.

4. Examples

4·1. *General*

We apply the above results to analyses of data taken from a study on determinants of cognitive development in young children, reported by R. Schumann in a dissertation of Johannes Gutenberg-University, Mainz, and a study of personality dispositions in high-school children (Kohlmann, Schumacher & Streit, 1987). We report only likelihood ratio test results without Bartlett corrections. The latter would only have resulted in higher p-values, implying an even better goodness-of-fit than we see already without this correction. Similarly, since no different conclusions were reached after looking at F-tests, we decided against reporting these. All the likelihood ratio test results and maximum likelihood estimates were computed by using a program described by Edwards (1989) for data from joint CG-distributions.

4·2. *Example with two continuous and one discrete variable*

Schumann was interested in how small children perform in tasks which require the simultaneous storage and transformation of information. The chosen tasks for 55 children in Kindergarten consisted of identifying among several alternatives the correct match to a standard picture. The total number of correct matches in 17 such tasks was taken as a quantitative measure of performance. It was considered to be the response variable, X, to a quantitative as well as a qualitative influence. The number of times the child looked from the standard to the alternatives or back, divided by 17, was taken as a quantitative measure for information gathering behaviour, Y, which is one particular type of problem solving strategy in these tasks. Furthermore, a qualitative variable with six categories was used to capture certain levels of the child's information processing capacity, A. This variable was regarded as a potential influence to information gathering behaviour, as well. Summary statistics are displayed in Table 1.

After assuming conditional normal distributions for X given A and Y and for Y given A the likelihood ratio test results in Table 2 were obtained. They show that hypotheses of constant residual variances, $\tau_{xx.y}(i)=\tau_{xx.y}$, of homogeneous regression coefficients of X given Y and A, $\beta_{xy}(i)=\beta_{xy}$, as well as constant residual variances of Y given A, $\sigma_{yy}(i)=\sigma_{yy}$, are well compatible with the observations. The estimates computed under the combination of these three hypotheses show small deviations from the observed summary statistics except in situations with rather low numbers of observations; see Table 3. The linear dependence of X and Y can thus be taken as being equal for all levels of A, the regression coefficients are estimated as $\beta_{xy}(i)=0{\cdot}46$ for all i. Nevertheless, A has a moderating effect on this dependence; it cannot be correctly evaluated in the marginal distribution of X and Y alone, where the regression coefficient is estimated as $\tilde{\beta}_{xy}=0{\cdot}76$ with a standard error of 0·15. This follows with Proposition 2, the good fit of the data to $Y \perp\!\!\!\perp A \mid X$, that is to $\alpha_{yx}(i)=\alpha_{yx}$, and with the poor fit of the data to $X \perp\!\!\!\perp Y \mid A$, that is to $\beta_{xy}=0$, and to $X \perp\!\!\!\perp A \mid Y$, that is to $\alpha_{xy}(i)=\alpha_{xy}$.

To summarize: the variability of performance, $\sigma_{xx}(i)$, the variability of information gathering behaviour, $\sigma_{yy}(i)$, and the increase in performance due to this particular problem

Table 1. *Summary statistics for the performance data*

Levels of A	Nos. of obs.	Means X	Means Y	St. dev. X	St. dev. Y	Corr. coeff. for (X, Y)	Regr. lines: X on Y Intercepts	Regr. lines: X on Y Slopes
1	14	5·86	4·99	2·21	2·23	0·57	3·0	0·56
2	4	9·14	5·64	2·25	2·49	0·45	6·9	0·40
3	4	11·75	6·37	1·50	2·26	0·56	9·4	0·37
4	4	7·75	6·56	2·63	1·76	0·60	1·9	0·89
5	5	7·20	6·06	3·56	2·55	0·53	2·7	0·74
6	14	13·00	7·92	2·04	3·11	0·54	10·2	0·36

A, information processing capacity; X, performance; Y, information gathering behaviour.

Table 2. *Selected likelihood ratio test results for the data in Table* 1

Hypothesis Symbol	Hypothesis Meaning	χ^2	Degrees of freedom	Fractile or p-value
H_1	$\tau_{xx.y}(i)=\tau_{xx.y}$	4·05	5	0·54
H_2	$\beta_{xy}(i)=\beta_{xy}$ given H_1	1·81	5	0·88
H_3	$\sigma_{yy}(i)=\sigma_{yy}$	3·58	5	0·61
H_4	$Y \perp\!\!\!\perp A \mid X$ given H_1, H_2, H_3	4·23	5	0·52
H_5	$X \perp\!\!\!\perp A \mid Y$ given H_1, H_2, H_3	45·92	5	<0·001
H_6	$X \perp\!\!\!\perp Y \mid A$ given H_1, H_2, H_3	16·27	1	<0·001

Table 3. *Estimated parameters after assuming hypotheses H_1, H_2 and H_3 defined in Table 2 to be satisfied*

Levels i of A	Nos. per cell i	Means X	Means Y	St. dev. X	St. dev. Y	Corr. coeff. of (X, Y)	Regr. lines: X on Y Intercepts	Slopes
1	14	5·86	4·99	2·20	2·44	0·506	3·6	0·457
2	14	9·14	5·64				6·6	
3	4	11·75	6·37				8·9	
4	4	7·75	6·56				4·8	
5	5	7·20	6·06				4·4	
6	14	13·00	7·92				9·4	

solving strategy, $\beta_{xy}(i)$, can be taken to be the same whatever the level, i, of the child's information processing capacity. This constant impact of information gathering behaviour on performance is strong, $p<0{\cdot}001$ for $X \perp\!\!\!\perp Y \mid A$, and the average level of performance corrected for this impact, $\alpha_{xy}(i)$, depends strongly on the levels of information processing capacity, with $p<0{\cdot}001$ for $X \perp\!\!\!\perp A \mid Y$. However, this capacity, A, has no, or at least a negligible, moderating effect on predicting information gathering behaviour, Y, from known values for performance, X, since $p=0{\cdot}52$ for $Y \perp\!\!\!\perp A \mid X$. This last result is directly relevant for understanding the type of dependence of X on Y and A. Since each of the influences (Y, A) provides a substantial contribution to predicting performance in addition to the other variable, it follows with $Y \perp\!\!\!\perp A \mid X$ that they are marginally associated influences and that each of them, alone, is also an important explanatory variable. As a consequence, no further tests are needed to decide via Proposition 2 that information processing capacity moderates the constant linear dependence of performance on information gathering behaviour. Thus, the study has established the relevance of this capacity of the memory as a background variable when studying effects of the chosen particular problem solving strategy on performance of preschool children.

4·3. *Example with one continuous and two discrete variables*

For the second set of data, the interplay of effects of two qualitative variables on a quantitative response is of main interest, but one wants to decide first whether estimates for this dependence can be obtained with or without pooling over the levels of a background variable. To decide on this we treat the two discrete influences as a single one, B.

The data are taken from research on effects of childrearing styles on the manifestation of anxiety as disposition in the child. Quantitative measures for anxiety and different educational styles were obtained as sum scores from questionnaires containing 20 items for the former (Spielberger, Gorsuch & Lushene, 1970) and 15 items for each of the different educational styles (Krohne, Kohlmann & Leidig, 1986). Anxiety is considered as a response to inconsistent behaviour of the parents. Data are available for 59 girls and 62 boys aged from 10 to 14 years, who all conceived of both parents as behaving not very supportively. The analysis is for our purposes restricted to this particular homogeneous subgroup of children, since effects of inconsistency on anxiety were expected to be similar and strongest under this condition.

In Table 4 summary statistics are reported for anxiety, X, as it was observed under four conditions describing distinct patterns of inconsistent behaviour of the parents, B,

Table 4. *Summary statistics for the anxiety data; A, sex of child; B, inconsistent behaviour of parents; n_{ij}, number of observations; X, anxiety in child*

	B			X		Estimated effects			
	Inconsistency						from obs. values	under	from obs. values collapsed
A	mother	father	n_{ij}	Mean	St. dev	name	given A	H_2, H_3	over A
Female	low	low	26	27·04	4·46	γ_0	30·61	30·76	30·56
		high	3	29·33	4·93	γ_1^A	1·05	1·13	—
	high	low	8	34·13	11·52	γ_1^B	−4·02	−4·22	−3·95
		high	22	36·14	8·23	γ_2^B	−1·59	−1·41	−1·66
Male	low	low	23	26·13	5·81	γ_3^B	1·90	1·74	1·94
		high	7	28·71	8·06	γ_{11}^{AB}	−0·60	0	—
	high	low	8	30·88	6·08	γ_{12}^{AB}	−0·74	0	—
		high	24	32·50	6·47	γ_{13}^{AB}	−0·57	0	—

Table 5. *Selected likelihood ratio results for the data in Table 4*

Hypothesis Symbol	Hypothesis Meaning	χ^2	Degrees of freedom	Fractile or p-value
H_1	$\pi_{i\|j} = \pi_i$	1·80	3	0·62
H_2	$\tau_{xx}(i,j) = \tau_{xx}$	16·92	7	0·02
H_3	$\gamma_{ij}^{AB} = 0$ given $\tau_{xx}(i,j) = \tau_{xx}$	1·23	3	0·75

for girls and boys, A, separately. The data indicate that inconsistency affects not only the level but also the variability of anxiety: the assumption of homogeneous residual variances of a normally distributed X given A and B is not well supported by the observations; see Table 5. We proceed, nevertheless, to estimate effects by assuming equal residual variances for the following reasons. First, we do not know, at present, how to obtain estimates of additive effects on the means if the residual variances are not taken to be constant; secondly, this assumption causes the maximum likelihood estimates of canonical parameters, but not of the means, to deviate from the observed ones; and thirdly, computation of marginal means from conditional ones does not involve variances; see Appendix 1.

Psychological theory did not predict an interactive effect of sex and inconsistency nor different probabilities for inconsistent behaviour of the parents for girls and boys. The observations are also far from indicating such effects; see Table 4. Consequently, one does, with Proposition 3, expect no moderating effect of sex on γ_j^B, the effects of parental inconsistency on anxiety in the child. One may therefore proceed by pooling the two samples of boys and girls to estimate the effects of parental behaviour on anxiety as disposition in the child; see Table 4.

5. CONCLUSIONS

Understanding conditions for the lack of a moderating effect is important for drawing conclusions from data. The above conditions help researchers to use their theoretical subject-matter expectations and knowledge from previous empirical investigations to make well-founded choices on such issues as the following.

Which variables need to be explicitly included into an analysis?
Can estimates of the directed measures of association be computed after pooling subgroups?
Are results of two studies, which differ by reporting one discrete influence consistent with each other?
Can the results of an empirical investigation give valid estimates of dependencies in the target population?

Clearly these questions are also of importance in more general settings than the ones discussed on this paper, like linear models with joint responses and several influences, response models with qualitative responses as logistic regressions or multinomial logit models or in situations with a potential quantitative moderator variable. Accordingly, more general results are needed.

Acknowledgement

I am extremely grateful to a referee, who found a serious error in a previous formulation of Propositions 1 and 2. Only after recognizing this error could I formulate Propositions 2 and 4.

Appendix 1

Relating marginal to conditional moments

The essential step in quantifying moderating effects, in (2·6) and (3·5) consists of nothing but computing marginal moments from conditional ones (Rao, 1973, p. 97) as

$$E(Z)=E_T E_{Z\mid T}(Z\mid t), \quad \operatorname{var}(Z)=E_T\{\operatorname{var}_{Z\mid T}(Z\mid t)\}+\operatorname{var}_T\{E_{Z\mid T}(Z\mid t)\}.$$

Starting from a slightly more general version of (2·4) by permitting the covariance matrix to depend on the levels of A, we have that the marginal moments are

$$E(Z)=[\tilde{\mu}_x \;\; \tilde{\mu}_y]=[\textstyle\sum_i \pi_i\mu_x(i) \;\; \sum_i \pi_i\mu_y(i)],$$

$$\tilde{\sigma}_{xy}=\textstyle\sum_i \pi_i\sigma_{xy}(i)+\sum_i \pi_i\{\mu_x(i)-\tilde{\mu}_x\}\{\mu_y(i)-\tilde{\mu}_y\},$$

$$\tilde{\sigma}_{yy}=\textstyle\sum_i \pi_i\sigma_{yy}(i)+\sum_i \pi_i\{\mu_y(i)-\tilde{\mu}_y\}^2.$$

With homogeneous regression coefficients for X on Y and A, that is $\beta_{xy}(i)=\beta_{xy}$, one gets from $\alpha_{xy}(i)=\mu_x(i)-\beta_{xy}\mu_y(i)$ and $\tilde{\alpha}_{xy}=\Sigma_i\pi_i\alpha_{xy}(i)$ that

$$\alpha_{xy}(i)-\tilde{\alpha}_{xy}=\{\mu_x(i)-\tilde{\mu}_x\}-\beta_{xy}\{\mu_y(i)-\tilde{\mu}_y\}.$$

Together with $\sigma_{xy}(i)=\beta_{xy}\sigma_{yy}(i)$ this permits us to write

$$\tilde{\sigma}_{xy}=\beta_{xy}\tilde{\sigma}_{yy}+\textstyle\sum \pi_i\{\alpha_{xy}(i)-\tilde{\alpha}_{xy}\}\{\mu_y(i)-\tilde{\mu}_y\}$$

and (2·6) follows with $\tilde{\beta}_{xy}=\tilde{\sigma}_{xy}/\tilde{\sigma}_{yy}$.

Starting from a slightly more general version of (3·1) and (3·2) by permitting means (μ_{ij}) and residual variances, $\tau_{xx}(i,j)$, to depend on the level combinations of A and B, we obtain the means after marginalizing over A as $E(X\mid B=j)=\tilde{\mu}_j=\Sigma_i\pi_{i\mid j}\mu_{ij}$, with $\pi_{i\mid j}=\operatorname{pr}(A=i\mid B=j)$.

With only main effects of A and B on X, that is $\mu_{ij}=\gamma_0+\gamma_i^A+\gamma_j^B$, this implies that

$$\tilde{\mu}_j=\gamma_0+\gamma_j^B+\textstyle\sum_i \pi_{i\mid j}\gamma_i^A,$$

and with the marginal effects defined by $\tilde{\gamma}_j^B=\tilde{\mu}_j-\Sigma_j\tilde{\mu}_j/J$ equation (3·5) follows.

APPENDIX 2

Relating marginal to conditional interactions of discrete variables

Given that X, A and B follow a homogeneous CG-distribution we can write for the joint density g_{ABX}

$$\log g_{ABX} = (\lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}) + (\eta^X + \eta_i^{AX} + \eta_j^{BX} + \eta_{ij}^{ABX})x - \tfrac{1}{2}\psi^X x^2,$$

where the λ's, η's and ψ are called the discrete, linear and quadratic interactions, satisfying symmetric constraints to insure uniqueness. They relate to the parameters in (3·1) and (3·2) via

$$\psi^X = 1/\tau_{xx}, \quad \tau_{xx}(\eta^X + \eta_i^{AX} + \eta_j^{BX}) = \gamma_0 + \gamma_i^A + \gamma_j^B, \quad \eta_{ij}^{ABX} \equiv 0,$$

$$\lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} = \log \pi_{ij} - \tfrac{1}{2}\log(2\pi\tau_{xx}) - \tfrac{1}{2}(\gamma_0 + \gamma_i^A + \gamma_j^B)^2/\tau_{xx}.$$

Discrete marginal interactions are the usual log linear interaction parameters for π_{ij}. In particular, the marginal two-factor interaction $\tilde{\lambda}_{ij}^{AB}$ is known to be defined as

$$\tilde{\lambda}_{ij}^{AB} = \log \pi_{ij} - (\textstyle\sum_i \log \pi_{ij})/I - (\sum_j \log \pi_{ij})/J + (\sum_{i,j} \log \pi_{ij})/(IJ).$$

Direct computation gives the result in (3·3).

REFERENCES

BARON, R. M. & KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Person Social Psychol.* **51**, 1173-82.

COHEN, J. & COHEN, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd. ed. Hillsdale, N.J.: Erlbaum.

EDWARDS, D. (1989). Hierarchical interaction models. *J. R. Statist. Soc.* To appear.

KOHLMANN, C. W., SCHUMACHER, A. & STREIT, R. (1987). Parental child rearing behavior and the development of trait anxiety in children: support as a moderator variable? *Anxiety Res.* **1**, 53-64.

KROHNE, H. W., KOHLMANN, C. W. & LEIDIG, S. (1986). Erziehungsstildeterminanten kindlicher Ängstlichkeit, Kompetenzerwartungen und Kompetenzen. *Z.f. Entwicklungspsychol. Pädagogische Psychol.* **18**, 70-88.

LAURITZEN, S. L. & WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* To appear.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

SAUNDERS, D. R. (1956). Moderator variables in prediction. *Educ. Psychol. Meas.* **16**, 209-22.

SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *J. R. Statist. Soc.* B **13**, 238-41.

SNEDECOR, G. W. & COCHRAN, W. G. (1967). *Statistical Methods*, 6th ed. Ames: Iowa State University Press.

SPIELBERGER, C. D., GORSUCH, R. L. & LUSHENE, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, Calif.: Consulting Psychologists.

WERMUTH, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. R. Statist. Soc.* B **49**, 353-64.

WERMUTH, N. (1989). Moderating effects in multivariate normal distributions. *Methodika.* To appear.

WHITTEMORE, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. R. Statist. Soc.* B **40**, 328-40.

YAKOWITZ, S. J. & SPAGINS, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39**, 209-14.

YULE, G. U. (1900). On the association of attributes in statistics. *Phil. Trans. R. Soc.* A **194**, 257-319.

ZEDECK, S. (1971). Problems with the use of "moderator" variables. *Psychol. Bull.* **76**, 295-310.

[*Received August* 1987. *Revised July* 1988]