In: Encyclopedia of Statistical Sciences. S. Kotz, C. Read and D. Banks (eds). New York: Wiley, Second Update Volume, 284-300.

languages.

Goulden's career as a part-time teacher at the University of Manitoba continued until 1948, when he moved to Ottawa to become Chief of the Cereal Crops Division in what is now Agriculture Canada. This move was another turning point in Goulden's career; he became more absorbed in administrative duties and less involved in day-to-day research activities. In 1955 he was appointed Director of the Experimental Farms Service of Agriculture Canada and four years later was promoted to Assistant Deputy Minister with responsibility for the newly formed Research Branch of Agricultural Canada. As an administrator Goulden brought about several organizational improvements to the research arm of Agriculture Canada. He retired from the civil service in 1962 but continued to be active, designing several exhibits for "Man the Provider" for the international exhibition, Expo 67, held in Montreal.

Cyril Goulden was the recipient of many honors. He was an elected fellow of the Royal Society of Canada (1941), an elected fellow of the American Statistical Association* (1952), and an honorary member of the Statistical Society of Canada* (1981). In 1958 he served as president of the Biometrics Society*. He was awarded an honorary LL.D. from the University of Saskatchewan (1954) and an honorary D.Sc. from the University of Manitoba (1964). In 1970 he received an Outstanding Achievement Award from the University of Minnesota, and ten years later was named to the Canadian Agricultural Hall of Fame.

References

[1] Bartlett, M. S. (1939). Statistics in theory and practice. *Nature*, **144**, 799–800.

[2] Box, J. F. (1978). *R. A. Fisher: The Life of a Scientist.* Wiley, New York.

[3] Cochran, W. G. (1940). Review of *Methods of Statistical Analysis. J. R. Statist. Soc. A*, **103**, 250–251.

[4] Cochran, W. G. and Cox, G. M. (1950). *Experimental Designs.* Wiley, New York.

[6] Goulden, C. H. (1929). *Statistical Methods in Agronomic Research... a Report Presented at the Annual Conference of Plant Breeders, June, 1929.* Canadian Seed Growers' Association, Ottawa.

[7] Goulden, C. H. (1931). Modern methods of field experimentation. *Sci. Agric.*, **11**, 681–701.

[8] Goulden, C. H. (1936). *Methods of Statistical Analysis.* Burgess, Minneapolis.

[9] Goulden, C. H. (1937). *Methods of Statistical Analysis*, rev. ed. Burgess, Minneapolis.

[10] Goulden, C. H. (1937). Efficiency in field trials of pseudo-factorial incomplete randomized block methods. *Can. J. Res.*, **15**, 231–241.

[11] Goulden, C. H. (1939). *Methods of Statistical Analysis.* Wiley, New York.

[12] Goulden, C. H. (1944). A uniform method of analysis for square lattice experiments. *Sci. Agric.* **23**, 115–136.

[13] Goulden, C. H. (1952). *Methods of Statistical Analysis,* 2nd ed. Wiley, New York.

[14] Goulden, C. H. and Elders, A. T. (1926). A statistical study of the characters of wheat varieties influencing yield. *Sci. Agric.*, **6**, 337–345.

[15] Goulden, C. H., Neatby, K. W., and Welsh, J. N. (1928). The inheritance of resistance to *Puccinia graminis tritici* in a cross between two varieties of *Triticum vulgare. Phytopathology*, **18**, 631–658.

[16] Johnson, T. (1967). The Dominion Rust Research Laboratory, Winnipeg, Manitoba, 1925–1957. Unpublished report seen courtesy of the Canadian Agriculture Research Station, Winnipeg.

[17] Kirk, L. E. and Goulden, C. H. (1925). Some statistical observations on a yield test of potato varieties. *Sci. Agric.*, **6**, 89–97.

[18] Owen, A. R. G. (1953). Review of *Methods of Statistical Analysis*, 2nd ed. *J. R. Statist. Soc. A*, **126**, 204–205.

[19] Student (1926). Mathematics and agronomy. *J. Amer. Soc. Agron.*, **18**, 703–719.

[20] Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *J. Agric. Sci.*, **26**, 424–455.

David Bellhouse

GRAPHICAL MARKOV MODELS

Graphical Markov models provide a flexible tool for formulating, analyzing, and interpreting relations among many variables. The models combine and generalize at least three different concepts developed at the turn of the last century: using graphs, in which variables are represented by nodes, to characterize and study processes by which joint distributions may have been generated (Sewell Wright [60, 61]), simplifying a joint distribution with the help of conditional independences (Andrei Markov* [36]), and specifying associations only for those variables which are in some sense nearest neighbors* in a graph (Willard Gibbs [21]).

Graphical Markov models are used now in many different areas, such as in expert systems* (Pearl [41], Neapolitan [38], Spiegelhalter et al. [45]), in decision analysis (Oliver and Smith [39]), for extensions of the notion of probability (Almond [1]), for attempts to model causal relations (Spirtes et al. [47]), and in multivariate statistics to set out and derive properties of structures (Lauritzen [30], Whittaker [57]) or to explain and summarize observed relations (Cox and Wermuth [15], Edwards [19]). We emphasize here their usefulness in observational studies*, where data are obtained on a considerable number of variables for each individual under investigation, where the isolation of relations between these variables is of main concern, and where available subject-matter knowledge is to be integrated well into model formulation, analysis, and interpretation. We illustrate in particular how graphical Markov models can aid

1. in setting up a first ordering of the variables under study to reflect knowledge about response variables of primary and secondary interest, about one or more levels of intermediate variables and purely explanatory variables,
2. in specifying hypotheses resulting from previous investigations,
3. in providing an overview of the analyses to be carried out,
4. in summarizing and interpreting the completed analysis, and
5. in predicting results in related investigations involving the same variables or a selection of the same variables.

The last important feature can also be made available to a number of traditional statistical models if they can be viewed [56] as special cases of graphical Markov models.

Most of the general concepts are introduced in the next two sections with the help of two specific research problems. A brief historical view is given first.

The geneticist Sewell Wright used directed graphs, which he called path diagrams, to describe hypothesized linear generating processes; he suggested estimating corresponding path coefficients and judging the goodness of fit of the process by comparing observed marginal correlations with those he derived as implied by the hypothesized process (*see* PATH ANALYSIS). It was not until much later that his estimated coefficients were identified by Tukey [50] as least squares regression coefficients of variables standardized to have mean zero and variance one, and his goodness-of-fit criterion was derived by Wermuth [52] as Wilks' likelihood ratio test [58] for Gaussian variables, provided the process can be represented by what is now called a decomposable concentration graph. These tests were not yet improved by Bartlett's adjustment* [12]. The same type of path analysis models were propagated in econometrics by Wold [59] and extended to discrete variables by Goodman [24] to be used in sociology and political science.

The notion of having a generating process which admits a causal interpretation was given up in extensions of path analysis aimed at modeling proper joint responses: extensions to simultaneous equation models [26], to linear structural equations* [27] and to chain graph models [32]. Similarities and distinctions between the different approaches were derived much later [53].

The notion of conditional independence*, which had been used by the probabilist Markov to simplify seemingly complex processes, was studied in detail by Dawid [17] and related by Speed [43] and Darroch et al. [16] to the undirected graphs which Gibbs had used to determine the total energy in a system of particles such as atoms of a gas.

Measures of association in graphical Markov models depend on the type of variables involved in a system. For instance, for Gaussian variables they are linear regression or correlation coefficients, for discrete variables they are odds ratios* studied early by Yule [62] and Bartlett [5], and for mixed variables they are based on different measures discussed by Olkin and Tate [40] and Cox [10].

Models with several variables which are now seen as special cases of graphical Markov models were developed quite separately, for instance, as linear models for contingency tables* [6, 23, 11, 7] and as covariance selection models for Gaussian variables by Dempster [18]. Analogies between independence interpretations and between likelihood ratio tests in these two model classes were recognized later by Wermuth [51]. Similarly, linear structure models in covariances [2, 27] which admit an independence interpretation were integrated into graphical Markov models by Cox and Wermuth [13] only many years after they had first been proposed, as were many of the generalized linear models* introduced by McCullagh and Nelder [37].

Of special importance for sparse data are models which permit explicit maximum likelihood estimation of parameters and testing of goodness of fit by considering only subsets of variables. Such decompositions of joint distributions were first studied by Haberman [25] and Sundberg [48] for contingency tables, by Speed and Kiiveri [44] for covariance selection, and by Lauritzen and Wermuth [32] for models with both discrete and Gaussian variables. Efficient algorithms for deciding on this property of a model from its graph have been designed [49, 33, 34].

Models for the same set of variables may differ with respect to their defining parameters. Then they typically correspond to different graphical representations, but they may nevertheless imply the same set of independence statements. This agreement has been termed *independence equivalence* of models and is one important aspect of the stronger requirement of equivalence in distributions. Early first results by Wermuth and Lauritzen [52, 55, 32] were generalized by Frydenberg [20].

This last is one active research area. Another is to incorporate latent variables into the model formulations and to integrate time series and survival analysis*. Also, further criteria are developed to read off a graph which conditional independences and which conditional associations are implied. Some of the available criteria are described here in the next-to-last section.

AN INTERVENTION STUDY

A first illustration of the use of graphical Markov models in observational studies is an intervention study involving 85 chronic pain patients and observations before and after three weeks of stationary treatment [35, 15]. There were three main objectives. We wanted to see whether results reported in earlier studies can be replicated, whether it is worthwhile to study measures of well-being of a chronic pain patient other than self-reported treatment success, and whether a psychotherapeutic placebo intervention shows sizeable effects.

This placebo intervention, offered to about one-third of the patients, consisted of telephone contacts, solely involving general information and interest in the patient's well-being, offered by one physician over a period of three months after stationary treatment, together with the option for the patient of phoning this physician again.

The available knowledge about the variables under study and decisions about potential relevant explanatory variables lead to the ordering of the variables in Fig. 1. Such an ordering is called a *dependence chain*, because the variables are arranged in sequence, in a chain of boxes. One variable is binary, the additional intervention, A; all 25 other variables are quantitative measurements.

There are three response variables of primary interest, measuring different aspects of the patient's well-being three months after the patient has left the hospital: the patient's own judgement of success of the stationary treatment, Y; the self-reported typical type of pain, Z_a; and a depression score, X_a. Whenever there are several variables that we want to treat on an equal footing in the sense of not specifying one variable as a response and another as an explanatory variable, then we list them in the same

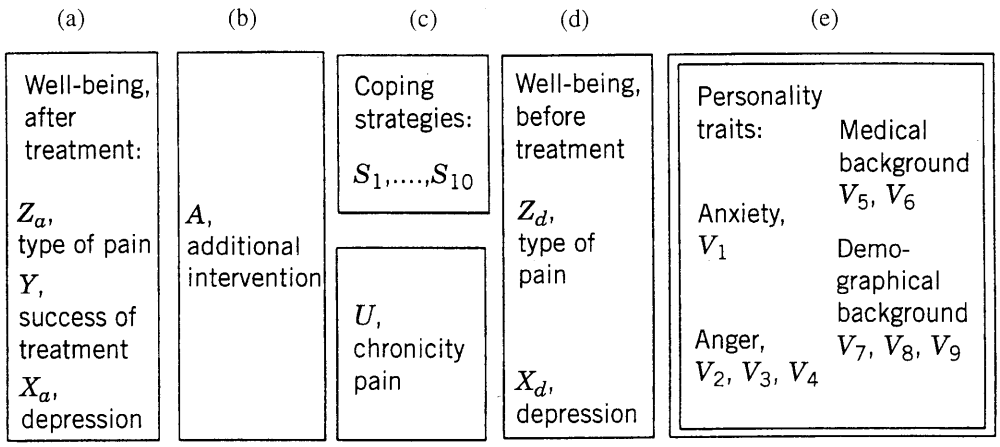


Figure 1 First ordering of types of variable for chronic-pain patients: patient's well-being three weeks after stationary treatment (box a) and before (box d); an intervention (box b) intermediate for the former and response to the latter. The stage of chronic pain, U, and the coping strategies S_i are stacked to display hypothesized independence given pretreatment well-being (box d), personality characteristics, and medical and demographic background variables (box e). The double-lined box indicates relations taken as given.

box. This is to imply that we investigate aspects of their joint (conditional) distribution given all variables listed in boxes to the right.

Several variables are *intermediate*, because they are taken as potentially explanatory for some variables in the system (listed in boxes to the left) and as response to others (listed in boxes to the right). Depression before treatment, X_d, for instance, is considered here as explanatory for chronicity of pain, U, but as a response to certain patient characteristics, $V_1, \ldots, V_9$, and, as another measure of well-being of the patient before treatment, on an equal footing with typical intensity of pain, Z_d.

In a previous investigation it had been claimed that certain strategies to cope with pain $S_1, \ldots, S_{10}$ are independent of chronicity of pain, U, given pretreatment conditions (Z_d, X_d), personality characteristics, and medical and demographic information about the patient $V_1, \ldots, V_9$. The two stacked boxes indicate this hypothesized conditional independence.

The set of variables to be taken as purely explanatory is listed to the far right end of the sequence, in a doubly lined box, to indicate that their relations are taken as given without being analyzed in much detail.

We illustrate two typical steps of a full analysis for these data: checks for possible interactive effects and plots showing the direction and size of some of the estimated conditional interactions. For a full analysis of joint responses a decision has to be made whether other responses on an equal footing are to be included in regression equations as additional regressors or not. If prediction of responses is an aim, the latter appears often most appropriate, and it leads to what we call a *multivariate regression* approach.* In this, each of the joint responses is taken separately in turn, to decide which of the potentially important explanatory variables are in fact directly important. Next it is checked whether any important symmetric associations remain among the responses after the regressions. To summarize such analyses, relations to important regressors enter as arrows between boxes and important symmetric associations as lines within boxes to give what is called in its most general form a *joint response chain graph.*

Plots to check systematically for nonlinear relations have recently been proposed by Cox and Wermuth [14]. Two such plots, displaying t-statistics for interactive effects from trivariate marginal distributions, are shown in Fig. 2. Most t-statistics, even if they are considerably larger than 2, lie along the line of unit slope in Fig. 2a and may therefore be interpreted as being well compatible with random variation. A few statistics deviate from the unit line. They indicate the presence of interactions, which are to be identified. Figure 2b shows that these marginal interactions do not involve chronicity of pain, U, coping strategies, or earlier variables (listed in boxes c, d, e), since the largest t-

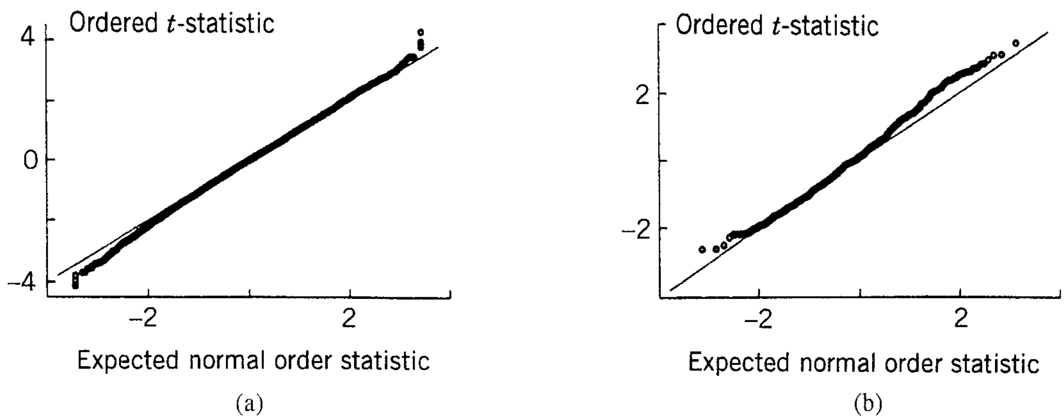


Figure 2 Normal probability plot of t-statistics of cross-product terms in trivariate regressions for variables on chronic pain. (a) Plot for all variables. (b) Plot for stage of chronic pain, U, coping strategies S_i, and their potential explanatory variables; each triple includes U.

values in this plot are small, that is, they are in absolute value near 2.

Two estimated interactive effects are displayed in Fig. 3. All other directly important explanatory variables are there kept at their average level, that is, mean values are inserted for all remaining regressors in the selected regression equation. Figure 3a shows that the psychotherapeutic placebo intervention, A, has essentially a positive effect on reported treatment success for patients with low pain intensity $\bar{z}_d - s_z$, but a negative effect for patients with typically high intensity of pain. Figure 3b shows that the hypothesized conditional independence between stage of chronic pain and strategies to cope has to be rejected: the risk for higher chronicity decreases with increasing use of one of these strategies by patients with low pain intensity, but it increases with use of the same strategy by patients with high pain intensity.

In this study the additional intervention is important for each of the three responses, the particular effect type depending on specific characteristics of the patient and on the coping strategies of the patient. The estimated relations, however, appear to be fairly complex relative to the sample size. For instance, for the self-reported treatment success there are three sizable interactive effects, and for typical pain intensity there are two alternative regression equations, which fit the data equally well and which both permit plausible interpretations. Hence it appears best to try to replicate some of the results with a reanalysis of previous data and in a study with more and different patients, instead of attempting to summarize results with a graph.

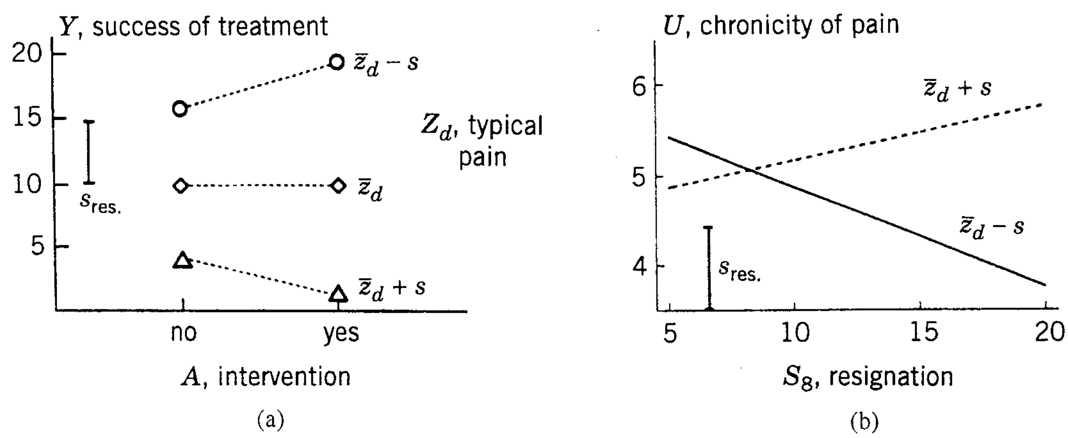


Figure 3 Plots to interpret interactive effects on two responses in treatment of chronic-pain data. Remaining directly explanatory variables are fixed at level 1 or at mean. (a) Response is Y, (b) response is U. Standard deviation of residuals in the regression is denoted by s_{res}; observed mean and standard deviation of a variable X are denoted by $\bar{x}$, s, respectively.

A COHORT STUDY

The second example is a cohort study [22, 15], in which the main aim is to identify important developments in a school and a student career which might increase the risk that a student stops studying without having received a degree.

Complete records were available for 2339 high school students on their average performance in high school and on questions regarding school career and demographic background. During their first year at university, responses to psychological questionnaires were obtained, and it was finally recorded whether students successfully completed their studies or dropped out of university.

Figure 4 shows the first ordering of the observed variables, based largely on the time sequence involved; it implies that relations between the variables are to be studied in five conditional distributions. For the variable in box *a*, all other variables are taken as potentially explanatory. The variables in box *b* are treated on an equal footing and are considered conditionally given those in boxes *c*, *d*, *e*, *f*. For the variable in box *c*, those in boxes *d*, *e*, *f* may be explanatory, but not those listed in boxes *a*, *b*, and so on. Finally, the variables in box *f* are treated as purely explanatory with associations not to be further specified by any model.

Checks for nonlinearities gave no indication of any strong interactions or nonlinear dependences among the ten variables, of which three are quantitative measurements (Y, X, Z) and seven are binary variables ($A, B, \ldots, F$). The procedure followed to study the relations among the variables via separate logistic and linear regression, as well as log-linear contingency table analyses, is summarized in the graphs of Figs. 5 and 6.

In studying the relations among the joint responses, X, Y, Z, we did not intend to identify good predictors, so that a *block regression* approach has been used. In it the other components of a response vector serve as additional regressors for each component considered in turn as a single response. This leads typically to simpler *independence structures*, that is, to larger sets of independences implied by a system. The reason is that fewer of the potentially important regressors become directly important if information on other responses on an equal footing is available in a regression equation.

Even with as many as ten variables, it would be possible to summarize the results of the analyses with a single chain graph, but we present separate graphs for the university variables given earlier variables and for the high school variables given the demographic background information in Figs. 7 and 8.

This not only is clearer, but illustrates one of the features of chain graph representations: separate analyses may be integrated into larger graphs in different ways to emphasize special

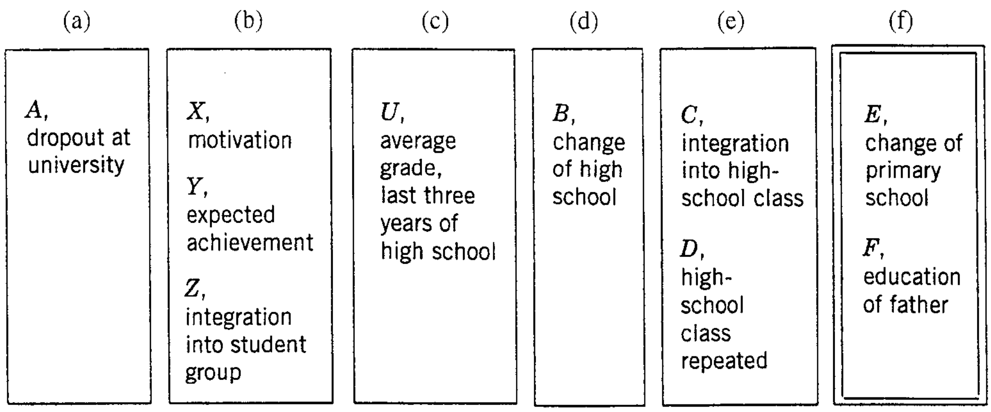
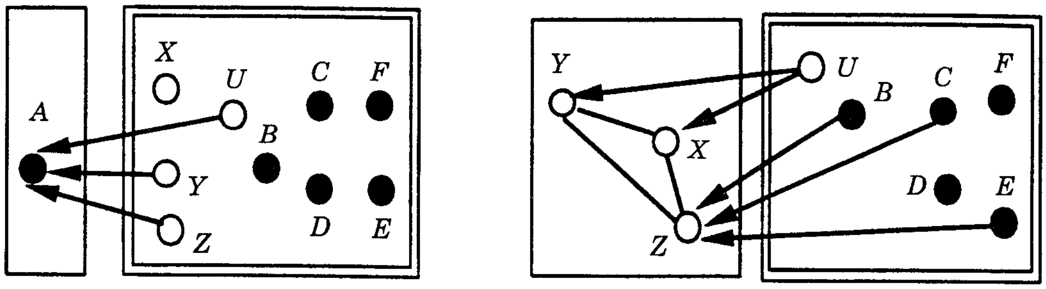


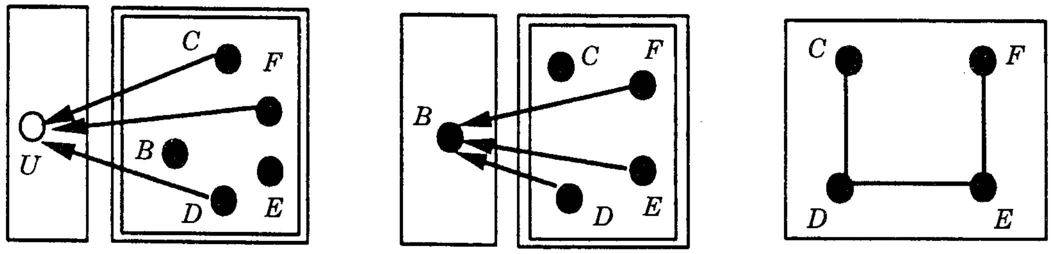
Figure 4 First ordering of variables for university students. Variable A, dropout at university (box a), is the response variable of primary interest; for instance, B, change of high school (box d), is an intermediate variable, potentially explanatory for dropping out at university (box a), for the student's attitudes towards his study situation (box b), and for grades at high school (box c), and also a potential response to other school-career and demographic variables (boxes e, f). Variables are treated on an equal footing in boxes b, e because no direction of dependence is specified—in box f because variables are purely explanatory.



Selected regression models

A: *Y*+*Z*+*U* *Y*: *X*+*Z*+*U*, *X*: *Y*+*Z*+*U*, *Z*: *Y*+*X*+*B*+*C*+*E*

Figure 5 Regression graphs showing results of analyses for university variables in cohort study.



Selected regressions Selected log-linear model

U: *C*+*D*+*F* *B*: *D*+*E*+*F* *CD*,*DE*,*EF*

Figure 6 Regression graphs and concentration graph showing results of analyses for high school variables in cohort study.

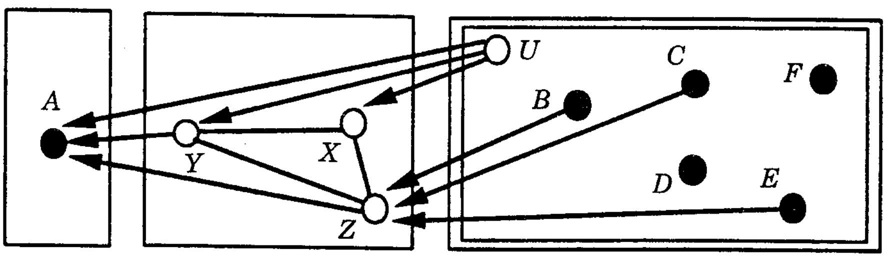


Figure 7 Independence graph for university variables given high school and background variables.

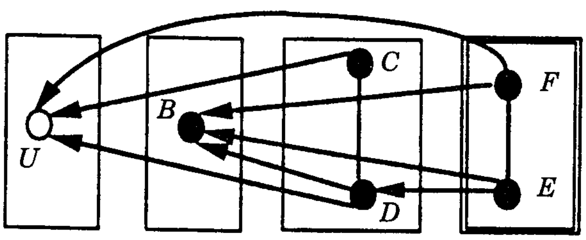


Figure 8 Independence graph for high-school variables given demographic background variables.

aspects, here the university components and the high school career.

A short qualitative interpretation of the graphs is that dropout, *A*, is directly influenced by the attitudinal variables *Y*, *Z* and by academic performance in the final years of high school as measured by the average mark, *U*, and that these variables in turn depend on some of the earlier properties.

A next step is to interpret the regression results more quantitatively, in particular the logistic regression* for *A*. One way of appreciating the sizes of the individual logistic regression coefficients is via Fig. 9, which shows the estimated probability for the three important explanatory variables, again in each plot holding the other two variables fixed at their means.

It is a matter of judgement when and how to introduce directed relations between the components of the psychological questionnaire scores *X*, *Y*, *Z*, but, partly because *X* is not directly explanatory for the primary response *A* and partly because of the simplification achieved, we have

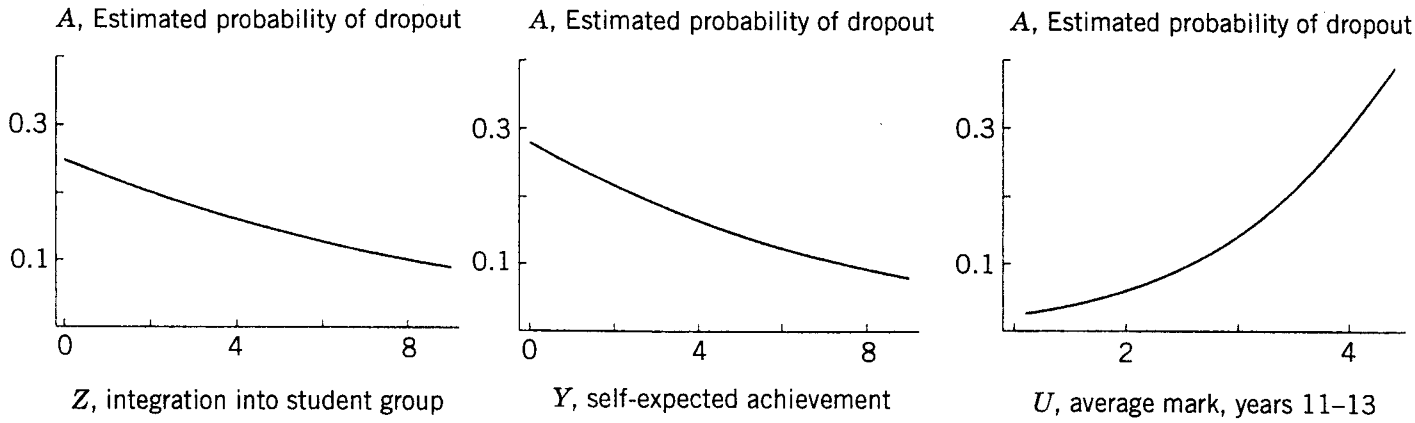


Figure 9 Fitted probability of university dropout, A, versus the directly explanatory variables in turn. In each plot two other variables are fixed at the mean; lower marks correspond to higher achievement.

explored treating Y as a response to Z and to U, ignoring X, that is, marginalizing over it. Similarly, we have marginalized over the variable C. This leads to the chain of Fig. 10, which shows only univariate responses.

An indirect path like the path from E to D to B to Z to A may be interpreted as follows: change of primary school, E, increases the risk that a high school class will have to be repeated, D, which in turn increases the risk that the student will change high school at least once during his school career, B. Once a high school change has been experienced, it becomes less likely that a student integrates well into his later student group (Z), and this in turn is a direct risk factor for A, leaving university without having obtained a degree. The overall effect of such a path may be moderate, however, if some of the relations along such a path are of moderate strength. This does not show in the graph, but is the case here for the edge DE.

The possibility of using the graph to infer additional independences and induced dependences is an additional appealing feature of graphical Markov models. We describe these in more detail for univariate-recursive regression graphs later, but give some formal definitions next.

SOME GRAPH TERMINOLOGY

For graphical Markov models, p *nodes* $V = \{1, \ldots, p\}$ in a graph denote random variables $Y_1, \ldots, Y_p$. There is at most one *edge* i, j between each pair of nodes i and j. Edges represent conditional association parameters in the distribution of Y_V. An edge may be directed and then drawn as an *arrow*, or it may be undirected and then drawn as a *line*. If a graph has only lines, then it is an *undirected graph*. The graph is *fully directed* if all its edges are arrows, and *partially directed* if some of its edges are lines and some are arrows.

An edge i, j has *no orientation* if it is a line; it has one of two possible *orientations* if it is an arrow, either pointing from j to i or pointing from i to j. A *path* of length $n - 1$ is a sequence of nodes $(i_1, \ldots, i_n)$ with successive edges (i_r, i_{r+1}) present in the graph; this is irrespective of the orientation of the edges. The graph obtained from any given one by ignor-

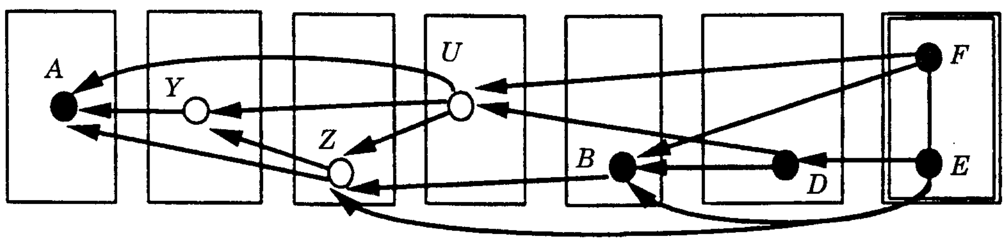


Figure 10 Independence graph for subset of variables concerning university dropout. Consistent with the results of Figs. 7 and 8 after marginalizing over variables X and C.

ing type and orientation of edges is called its *skeleton*.

Two nodes i, j are said to be *adjacent* or *neighbors* if they are connected by an edge, and they have a *common neighbor* t if t is adjacent to both i and j. Three types of common neighbor nodes t of i, j can be distinguished in a directed path as shown in Fig. 11*a* to *c*. Two arrows point to i and j from a *source node* t (Fig. 11*a*); a *transition node* t has one incoming arrow, say from j, and one outgoing arrow (Fig. 11*b*); and a *sink node* t has two arrows pointing at it from each of i and j (Fig. 11*c*). Because two arrows meet head on at a sink node, it is also called a *collision node*.

A graph constructed from a given one by keeping nodes and edges present within a selected subset S of nodes is an *induced subgraph*. If the induced subgraph of three nodes i, t, j has exactly two edges, it is a V-*configuration*, and if it is one of the paths of Fig. 11*a*, *b*, *c*, it is *source*-, *transition*-, or *sink-oriented*, respectively. Similarly, a subgraph as in Fig. 11*d* is called a sink-oriented U-*configuration*.

A path containing a collision node is a *collision path*, and a path is said to be *collisionless* otherwise. A path of arrows leading to i from j via transition nodes is called a *direction-preserving* path. In such a path i is a *descendant* of node j, and node j is an *ancestor* of i.

A *cycle* is a path leading from a node back to itself and a *directed cycle* is a (partially) oriented cycle without a sink-oriented V- or U-configuration along it.

Induced subgraphs which are complete in the sense of having all nodes joined but which become incomplete if even one other node is added are the *cliques* of the graph. The set of cliques of a graph without arrows and only full lines, points to the set of minimal sufficient statistics for a corresponding exponential family model. This special role of nearest neighbors is closely related to simplifying log-linear contingency table analyses and to the Hammersley–Clifford theorem* [16].

JOINT-RESPONSE GRAPHS AND GAUSSIAN SYSTEMS

The arrangement of the variables for a *joint-response chain graph* is, here, into boxes from left to right, starting with responses of primary interest. There are in general several nodes within each box corresponding to joint responses, i.e., to variables to be considered on equal footing. This corresponds to an ordered partition of the set V of all p nodes into subsets as $V = (a, b, c, \ldots)$, to the dependence chain. Different types of edge [15] add flexibility in formulating distinct structures.

Arrows pointing to any one box are either all dashed or all full arrows. Dashed arrows to a node i indicate that regression of the single variable Y_i on variables in boxes to the right of i is considered, whereas full arrows mean that the regression is taken both on variables in boxes to the right of i and on the variables in the same box as i. In this way dashed arrows indicate a multivariate regression and full arrows point to a block regression approach.

Within a box there is either a full-line graph or a dashed-line graph, each considered conditionally given variables in boxes to the right. For instance, a dashed line for a pair (i, j) in box b means the presence of conditional association between Y_i and Y_j given $\mathbf{Y}_c, \ldots$; a full line for a pair (i, j) in box b means the presence of conditional association between Y_i and Y_j given $\mathbf{Y}_{b\setminus\{i,j\}}$, $\mathbf{Y}_c$, $\ldots$.

Block regressions are combined with full-line response boxes, while for a multivariate regression there is a choice between a full-

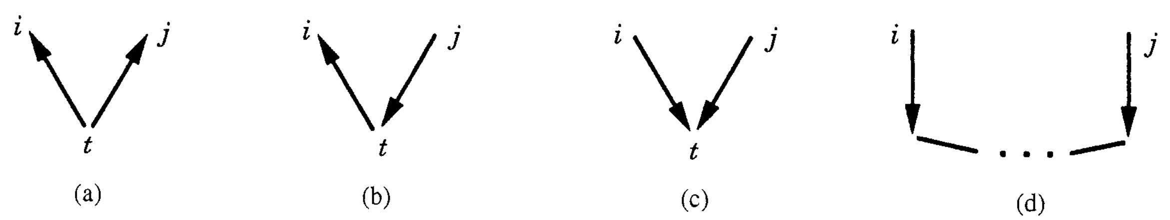


Figure 11 Some V-configurations with t (a) a source node, (b) a transition node, (c) a sink or collision node, (d) a sink-oriented U-configuration.

and a dashed-line graph for the corresponding joint responses. Figure 12 shows two joint-response graphs having the same skeleton but corresponding to different models.

To relate formally the corresponding parametrization in a joint Gaussian distribution it is convenient to think of the $p \times 1$ vector variable $\mathbf{Y}$ as having mean zero and being partitioned into component column vectors $\mathbf{Y}_a, \mathbf{Y}_b, \ldots,$ of dimensions $p_a \times 1$, $p_b \times 1, \ldots$.

For a dependence chain of three elements let the covariance matrix $\boldsymbol{\Sigma}$ and the concentration matrix $\boldsymbol{\Sigma}^{-1}$ be partitioned accordingly as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_{ac} \\ \cdot & \boldsymbol{\Sigma}_{bb} & \boldsymbol{\Sigma}_{bc} \\ \cdot & \cdot & \boldsymbol{\Sigma}_{cc} \end{pmatrix},$$

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{aa} & \boldsymbol{\Sigma}^{ab} & \boldsymbol{\Sigma}^{ac} \\ \cdot & \boldsymbol{\Sigma}^{bb} & \boldsymbol{\Sigma}^{bc} \\ \cdot & \cdot & \boldsymbol{\Sigma}^{cc} \end{pmatrix}.$$

A zero off-diagonal element in $\boldsymbol{\Sigma}$ specifies a marginal independence; a zero off-diagonal element in $\boldsymbol{\Sigma}^{-1}$ implies a conditional independence statement given all other remaining variables [51; 15, p. 69]. Independence statements with other conditioning sets arise as zero regression coefficients in different regressions as follows.

From regressions of each component of $\mathbf{Y}_b$ on $\mathbf{Y}_c$, regression coefficients are obtained (in the position of $\boldsymbol{\Sigma}_{bc}$ by sweeping $\boldsymbol{\Sigma}$ on c or resweeping $\boldsymbol{\Sigma}^{-1}$ on (a, b) in a $p_b \times p_c$ matrix $\mathbf{B}_{b|c}$:

$$\mathbf{B}_{b|c} = \boldsymbol{\Sigma}_{bc}\boldsymbol{\Sigma}_{cc}^{-1} = -(\boldsymbol{\Sigma}^{bb.a})^{-1}\boldsymbol{\Sigma}^{bc.a}.$$

In this each row corresponds to regression coefficients of $\mathbf{Y}_b$ in the regression of one of the components of $\mathbf{Y}_a$ on $\mathbf{Y}_b$; $\boldsymbol{\Sigma}^{bb.a}$, $\boldsymbol{\Sigma}^{bc.a}$ denote submatrices of the concentration matrix in the marginal joint distribution of $\mathbf{Y}_b$, $\mathbf{Y}_c$.

From regressions of each component of $\mathbf{Y}_a$ on both $\mathbf{Y}_b$ and $\mathbf{Y}_c$ we get (in the positions of $\boldsymbol{\Sigma}_{ab}$, $\boldsymbol{\Sigma}_{ac}$ by sweeping $\boldsymbol{\Sigma}$ on b, c or resweeping $\boldsymbol{\Sigma}^{-1}$ on a) the $p_a \times (p_b + p_c)$ matrix

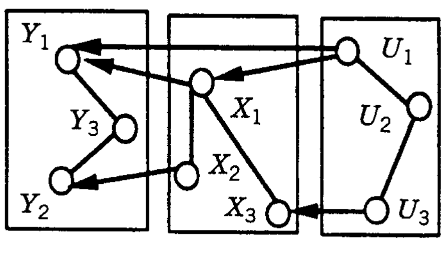
$$\begin{aligned}\mathbf{B}_{a|bc} &= (\mathbf{B}_{a|b.c}, \mathbf{B}_{a|c.b}) \\ &= (\boldsymbol{\Sigma}_{ab}, \boldsymbol{\Sigma}_{ac})\begin{pmatrix} \boldsymbol{\Sigma}_{bb} & \boldsymbol{\Sigma}_{bc} \\ \boldsymbol{\Sigma}_{cb} & \boldsymbol{\Sigma}_{cc} \end{pmatrix}^{-1} \\ &= -(\boldsymbol{\Sigma}^{aa})^{-1}(\boldsymbol{\Sigma}^{ab}, \boldsymbol{\Sigma}^{ac}).\end{aligned}$$

Then, the parameters of a pure multivariate-regression chain of only dashed edges result from the transformation $\mathbf{Z} = \mathbf{AY}$, and the parameters of a pure block-regression chain of only full edges, such as shown in Fig. 12*a*, result from the further transformation $\mathbf{X} = \mathbf{DZ}$, where $\mathbf{D} = \mathbf{T}^{-1}$ is a block-diagonal matrix containing concentration matrices of the distributions of $\mathbf{Y}_a$ given $\mathbf{Y}_b$ and $\mathbf{Y}_c$, of $\mathbf{Y}_b$ given $\mathbf{Y}_c$, and of $\mathbf{Y}_c$, and

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_{aa} & -\mathbf{B}_{a|b.c} & -\mathbf{B}_{a|c.b} \\ 0 & \mathbf{I}_{bb} & -\mathbf{B}_{b|c} \\ 0 & 0 & \mathbf{I}_{cc} \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa.bc} & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_{bb.c} & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{cc} \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} \boldsymbol{\Sigma}^{aa} & 0 & 0 \\ 0 & \boldsymbol{\Sigma}^{bb.a} & 0 \\ 0 & 0 & \boldsymbol{\Sigma}^{cc.ab} \end{pmatrix}.$$



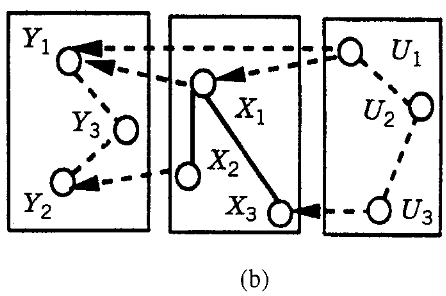


Figure 12. Two joint-response chain graphs with the same arrows and lines but different types of edges, implying different independence statements. For instance, absence of an edge for (Y_1, Y_2) means in (a) $Y_1 \perp\!\!\!\perp Y_2 | (Y_3, X_1, \ldots, U_3)$, in (b) $Y_1 \perp\!\!\!\perp Y_2 | (X_1, \ldots, U_3)$; absence of an edge for (X_1, U_3) means in (a) $X_1 \perp\!\!\!\perp U_3 | (X_2, X_3, U_1, U_2)$, in (b) $X_1 \perp\!\!\!\perp U_3 | (U_1, U_2)$; absence of an edge for (X_2, X_3) means $X_2 \perp\!\!\!\perp X_3 | (X_1, U_1, U_2, U_3)$ in both.

Here identity block matrices are denoted by **I**. The regression coefficients of a multivariate-regression chain are in **A**; those of a full block-regression chain are obtained from the matrix **DA** after dividing each row by the corresponding diagonal element [53]. The variance of **Z** is $\mathbf{A}^T\mathbf{\Sigma}\mathbf{A} = \mathbf{T}$; the variance of **X** is **D** and coincides with the diagonal block matrices of **DA**.

The parameters in a mixed dashed- and full-edge response graph coincide with those of a block-regression chain for full lines and with those of a multivariate regression chain for dashed edges. For instance, the model of Fig. 12*b* is like a multivariate regression chain, except that for the distribution of $\mathbf{Y}_b$ given $\mathbf{Y}_c$ the inverse covariance matrix is considered instead of the covariance matrix, i.e.,

$$\begin{aligned}\Sigma_{bb.c}^{-1} &= (\Sigma_{bb} - \Sigma_{bc}\Sigma_{cc}^{-1}\Sigma_{cb})^{-1} \\ &= \Sigma^{bb.a} = \Sigma^{bb} - \Sigma^{ba}(\Sigma^{aa})^{-1}\Sigma^{ab}.\end{aligned}$$

The transformation from covariance to concentration parameters introduces no complications, since mean and concentration parameters vary independently. This property holds more generally for mixed parametrizations with moment and canonical parameters in exponential families (Barndorff-Nielsen [4, p. 122]). For joint Gaussian distributions the numbers of parameters obtained with complete joint-response models, i.e., with unrestricted regression chains, coincide in the three different approaches described.

Full-edge joint-response graphs for discrete and continuous variables may correspond to joint conditional Gaussian distributions [32], in which null values of interaction parameters indicate the independences specified with the graph. With nonlinear and higher-order interactive effects permitted, the individual regressions specified with a graph capture independences in some non-Gaussian joint distribution.

Subclasses of joint-response chain graphs arise as follows. If the set of nodes, V, is not partitioned, so that there is just one box and no arrows, the dashed-line graph is a *covariance graph* and the full-line graph a *concentration graph*, because their edges correspond for a joint Gaussian distribution to elements in the overall covariance and concentration matrix, respectively. One main distinction is that concentration graphs with some edges missing point to sets of minimal sufficient statistics which simplify estimation, while this is not so for covariance graphs.

If V is partitioned into just two elements and the distribution of the explanatory variables is regarded as fixed by drawing a double-lined box for them, then the joint-response chain graph is reduced to a *regression graph*. Finally, if the partitioning is into single nodes, so that there are no joint responses (i.e., there are as many boxes as responses), then this is a *univariate-recursive regression graph* as in Fig. 10.

Some results are available on when two general joint-response graphs with identical skeletons imply the same set of independence statements. Such a condition for independence equivalence is that for any two full-edge graphs [20] the sink-oriented V- and U-configurations coincide. It explains, for instance, why the subgraph of nodes C, D, E, F in Fig. 8 has the same independence interpretation as the graph for these nodes in Fig. 6. For full-line, dashed-arrow graphs another criterion involving V- and U-configurations has been given by Anderson et al. [3].

Joint-response graphs arranged in boxes do not contain directed cycles. Graphs with directed cycles, i.e., *cyclic graphs*, were studied and related to some linear structural equations by Spirtes [46] and Koster [29]. In them an edge may be missing but not correspond to an independence statement.

TYPES OF FULLY DIRECTED INDEPENDENCE GRAPHS

In a *directed acyclic graph* G^V_{dag}, all edges are directed and there is no direction-preserving path from a node back to itself. Given the set of ancestors adjacent to any node i, this graph defines Y_i to be conditionally independent of the remaining ancestors of i.

In the contexts we are concerned with, the order of the variables is specified from subject-matter knowledge about the variables and, for each response in turn, by decisions about sets of variables considered to be potentially explanatory. A recursive ordering is indicated by drawing a chain of boxes around nodes 1, . . . ,p. The order need not be complete, in the sense that sets of several responses may be conditionally independent given all variables in boxes to the right; then the nodes of these sets are drawn in *stacked boxes*. If in Fig. 10 the edge E, F is oriented to point to E, then the fully directed graph obtained by deleting all boxes is directed acyclic.

Given such an order, we have a *univariate recursive regression system*, that is, a sequence of conditional distributions with Y_1 regressed on $Y_2, \ldots, Y_p$; Y_2 regressed on $Y_3, \ldots, Y_p$; and so on up to Y_{p-1} regressed on Y_p. Each *response* Y_i has then *potentially explanatory* variables $Y_{i+1}, \ldots, Y_p$, and we assume that its conditional dependence on Y_j given its remaining potentially explanatory variables can be captured by a set of parameters, the null values of which imply the corresponding independence statement, denoted by $Y_i \perp\!\!\!\perp Y_j | Y_{\{i+1,\ldots,p\}\setminus\{j\}}$, for $i < j$. That the independence structures of a directed acyclic graph and of a univariate-recursive regression graph coincide follows from what has been called the equivalence of local and pairwise Markov properties (Lauritzen et al. [31]).

The graphical representation of such a system, the univariate recursive regression graph, is a directed acyclic graph with two additional features: each edge present represents a specific nonvanishing conditional dependence, and each edge absent represents one particular conditional independence statement for the variable pair involved. We then say that the joint distribution is generated over the given G^V_{dag}, or that G^V_{dag} is a *generating graph*, because it is to represent a process by which the data could have been generated.

By considering G^V_{dag} as a generating graph, we mean that it is to have the same properties as a univariate recursive regression graph but without drawing the boxes; the order is then often indicated by a numbering of the nodes $(1, \ldots, p)$ whereby a set of r stacked nodes can obtain any one of the $r!$ numberings possible without affecting the independence structure of the system.

RELATIONS INDUCED BY DIRECTED ACYCLIC GENERATING GRAPHS

The possibility of reading all independence statements directly off a graph has been made available by so-called *separation criteria* for a graph. Such have been given to date for the two types of undirected graphs [16, 28], for partially directed full-line graphs [20, 8], and for directed acyclic graphs [41, 42, 31]. They permit one to conclude, under fairly general conditions on the joint distributions, that Y_A is independent of Y_B given Y_C, provided that the criterion is satisfied. The simplest criterion is for concentration graphs: C separates A from B in G^V_{con} for a nondegenerate joint distribution if every path from A to B has a node in C.

The separation criterion for directed acyclic graphs is more complex. To present it in a form which permits one to read induced associations directly off the graph, we define a path to be *active* whenever it is correlation-inducing in Gaussian systems. More precisely, in a joint Gaussian distribution generated over G^V_{dag} an active path between i and j relative to a set C introduces a nonzero component to $\rho_{ij.C}$ if the path is *stable*, that is, if correlations to each edge along it are strictly nonzero given C. Such edges may either be present in G^V_{dag} or be generated after conditioning on C.

A path between nodes i and j, $i \leqslant j$, in a directed acyclic graph G^V_{dag} is *active relative to* C if either

1. it is collisionless with every node along it outside C, or
2. a collisionless path wholly outside C is generated from it by completing with a line the nonadjacent nodes of every sink-oriented V-configuration having a descendant in C.

The definition is illustrated with Fig. 13.

Note that a path is collisionless if i is a descendant of j (Fig. 13a) or if there is a source node t which is an ancestor to both i and j (Fig. 13b). The contribution of such a path to the conditional association between i and j given C is obtained by marginalizing over all nodes along it. Similarly, the contribution of

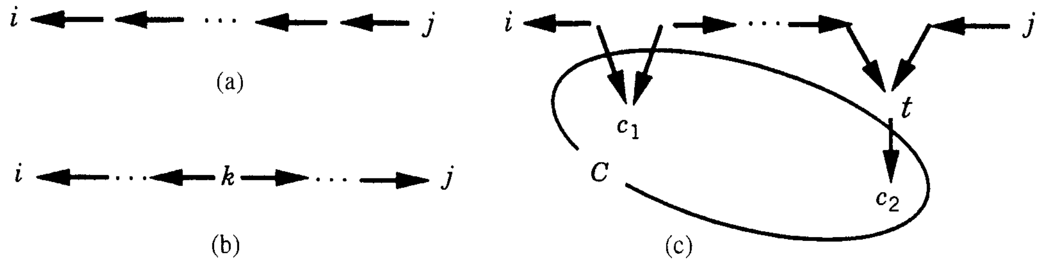


Figure 13 Active paths for a pair (i, j) relative to C: (a) collisionless active path with i descendant of j, (b) collisionless active path with i and j descendants of t (both have all nodes along it outside C), and (c) active collision path: a collisionless path outside C is generated from it by conditioning on C, i.e., a path not touching the collision nodes c_1 and t results after joining with a line the nonadjacent nodes in every sink-oriented V-configuration having a descendant in C.

an active collision path results by marginalizing over all nodes along the collisionless path, which only gets generated after conditioning on nodes in C.

Separation effects in directed acyclic graphs can be specified for disjoint subsets A, B, C of V as follows: $Y_A \perp\!\!\!\perp Y_B | Y_C$ if in G^V_{dag} there is no active path between A and B relative to C.

A stronger result of asserting dependence is possible whenever the joint distribution is nondegenerate and satisfies a condition which has been called *lack of parametric cancellation.* We stress that often representing dependences is as important as representing independences.

An association is induced by a generating graph G^V_{dag} for nondegenerate systems without parametric cancellation relative to C as follows: Y_i and Y_j are conditionally dependent given Y_C if in G^V_{dag} there is an active path between i and j relative to C.

The stronger result applies in essence to other than Gaussian systems provided they are *quasi-linear,* that is, any dependence present has a linear component such that the vanishing of the least squares regression coefficient implies or closely approximates an independence statement. Excluded thereby are situations in which dependences are so curved, or involve such a special high-order interaction, that they correspond to a vanishing correlation.

A *parametric cancellation* is a very special constellation among parameters such that an independence statement holds even though it is not implied by the generating graph, that is, even though it cannot be derived from the separation criterion. Thus the specific numerical values of the parameters are such that an independence arises that does not hold in general for structures associated with the given graph.

In a stable Gaussian system there is parametric cancellation only if the contributions to $\rho_{ij.C}$ of several active paths between i, j relative to C add up to give zero [54]. Examples are given here with the simple system of Fig. 14.

The graph of Fig. 14 specifies two independences: $X_1 \perp\!\!\!\perp X_4 | (X_2, X_3)$ and $X_2 \perp\!\!\!\perp X_3 | X_4$. The separation criterion tells us that the graph implies no marginal independence for the pair X_1, X_4 and no conditional independence for X_2, X_3 given X_1. Nevertheless, each of these additional independences may hold even in a stable Gaussian system whenever effects of different paths cancel each other. With the special constellation of correlation coefficients given by $\rho_{12} = \rho_{13} = \rho_{24} = \rho_{34} = \rho$, $\rho_{23} = \rho^2$, and $\rho_{14} = 2\rho^2/(1 + \rho^2)$, it follows that $\rho_{23.1} = 0$, because then the contribution to $\rho_{23.1}$ from the collisionless path $(2, 4, 3)$ and from the collision path $(2, 1, 3)$ cancel. If, instead, the correlation coefficients satisfy $\rho_{12}\rho_{24} = -\rho_{13}\rho_{34}$, it follows that $\rho_{14} = 0$, because the nonzero contribution to ρ_{14} of the path $(1, 2, 4)$ cancels the nonzero contribution of the path $(1, 3, 4)$.

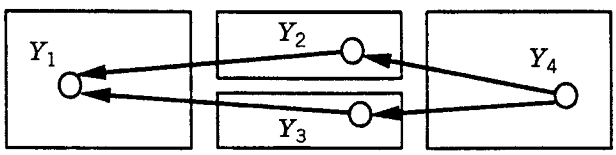


Figure 14 A simple stable Gaussian system in which effects of two paths may cancel: for ρ_{14} those of paths (1,2,4) and (1,3,4), and for $\rho_{23\cdot1}$ those of paths (2,1,3) and (2,4,3).

PREDICTION OF RESULTS IN RELATED INVESTIGATIONS

We now illustrate the use of the results stated in the preceding section for deriving consequences of a given hypothesized generating process.

The simple generating graph in five nodes of Fig. 15*a* consists just of one path, so that there can be no parametric cancellation in a corresponding stable system. This path is a collision path with node 1 being the single collision node in the system. Thus, in this graph a collision path between any pair of nodes i, j is active relative to C whenever C includes node 1. Then the generating graph does not imply $Y_i \perp\!\!\!\perp Y_j | Y_C$ for general joint distributions, and it implies Y_i dependent on Y_j given Y_C for a nondegenerate stable Gaussian system and for systems that behave similarly, such as the following system of main-effect regressions in mixed variables.

Let the joint distribution be generated in standardized variables Y, X, Z corresponding to nodes $1, 2, 5$ and in binary variables A, B, each taking values -1 and 1 with probability 0.5, such that the following set of regressions defines the joint distribution:

$$E\{Y|(X,A)\} = \rho_{yx}x + \rho_{ya}i\,,$$

$$\mathrm{var}\{Y|(X,A)\} = 1 - \rho_{yx}^2 - \rho_{ya}^2\,,$$

$$\mathrm{logit}\left(\pi_{1|x}^{A|X}\right) = \log \frac{\pi_{1|x}^{A|X}}{\pi_{-1|x}^{A|X}} = \frac{2\rho_{ax}x}{1 - \rho_{ax}^2}\,,$$

$$\mathrm{logit}\left(\pi_{1|x}^{B|X}\right) = \frac{2\rho_{bx}x}{1 - \rho_{bx}^2}\,.$$

Here ρ denotes a strictly nonzero correlation coefficient and, e.g., $\pi_{1|x}^{A|X}$ the conditional probability that A takes value 1 given $X = x$.

Then the joint distribution has as generating graph G_{dag}^V of Fig. 15*a*, and it is quasilinear.

If we choose C to be the empty set and draw a dashed-line edge for any pair for which marginal independence is not implied, the overall *induced covariance graph* of Fig. 15*c* is obtained, G_{cov}^V. Similarly, if for each i, j we take C to be the set of all remaining nodes ($C = V\backslash\{i,j\}$), and draw a full-line edge whenever $Y_i \perp\!\!\!\perp Y_j | Y_{V\backslash\{i,j\}}$ is not implied, then the overall *induced concentration graph* G_{con}^V of Fig. 15*c* results.

The following construction criteria for these graphs can be formulated. A given generating graph G_{dag}^V induces an edge i, j in the overall covariance graph G_{cov}^V if and only if in the generating graph there is a collisionless path between the two nodes, and it induces an edge i, j in the overall concentration graph G_{con}^V if and only if in the generating graph either i, j is an edge or nodes i and j have a common collision node.

More generally, for any selected subset S of all nodes V (where S includes i, j) a generating graph G_{dag}^V induces a dashed-line edge in the conditional covariance graph $G_{\mathrm{cov}}^{S.C}$ for nodes i, j if and only if in the generating graph there is an active path between i and j relative to C and it induces a full-line edge in the conditional concentration graph $G_{\mathrm{cov}}^{S.C}$ for nodes i, j if and only if in the generating graph either i, j is an edge or there is an active path between i and j relative to $S \cup C$ (Wermuth and Cox [54]).

Figure 16 shows a conditional covariance graph and a marginal concentration graph induced by the generating graph of Fig. 15*a*. These are not induced subgraphs of the corresponding overall graphs. For a stable Gaussian

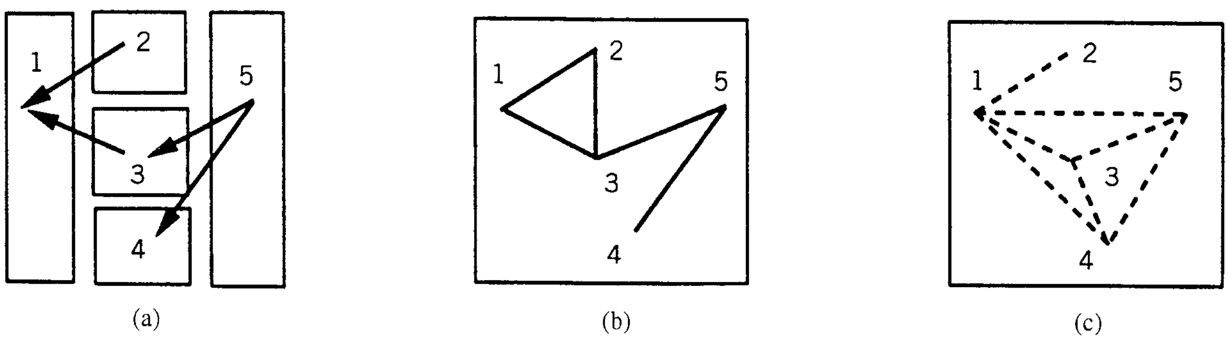


Figure 15 (*a*) A generating graph G_{dag}^V with five nodes; (*b*) the overall concentration graph G_{con}^V induced by it; and (*c*) the overall covariance graph G_{cov}^V induced by it.

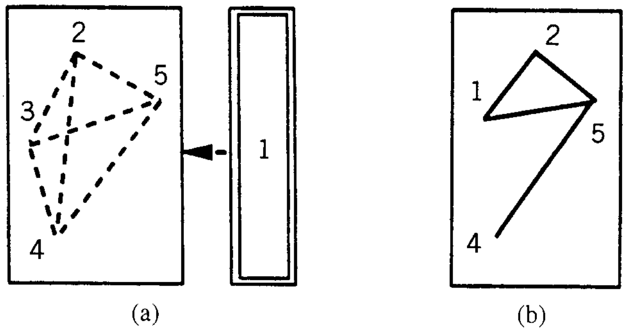


Figure 16 Further graphs induced by the generating graph of Fig. 15*a*: (*a*) the covariance graph $G_{\text{cov}}^{S.C}$ with $S = \{2,3,4,5\}$ given $C = \{1\}$, and (*b*) the concentration graph G_{con}^{S} with $S = \{1,2,4,5\}$.

system they show the zero and nonzero entries in $\Sigma_{SS.C}$ with $S = \{2,3,4,5\}$, $C = \{1\}$ and in Σ_{SS}^{-1} with $S = \{1,2,4,5\}$.

These results may be used to predict implications of a hypothesized generating system for different analyses with the same set of data. Prediction of results in related investigations may be illustrated with the help of Fig. 10, which summarizes some aspects of the study in the section "A Cohort Study."

Suppose, for instance, there is a study of academics, that is, of persons who have received a university degree, so that we condition on level 1 of variable A. Suppose further that information on the other seven variables shown in Fig. 10 is available. Then we may predict (for instance, with the induced covariance graph of seven nodes given A) which variable pairs should be marginally independent in this subpopulation, provided the graph of Fig. 10 is an adequate description of how the data are generated.

Suppose as a further example for A at level 1, we consider the relation between U and B given D, E, F. Figure 10 specifies $U \perp\!\!\!\perp B|(D,F)$, but an active path between U and B via Z relative to A. This implies, for distributions that are like stable Gaussian systems without cancellation of path effects, that U and B are dependent given A, D, F. Indeed, strong conditional association is typically obtained between U and B given D, F in similar studies of German academics.

Such possibilities of deriving consequences implied by a particular model mean that the important general principle of making a hypothesis elaborate, discussed in detail by Cochran [9], can be applied to these multivariate structures.

References

[1] Almond, R. (1995). *Graphical Belief Modelling*. Chapman & Hall, London.

[2] Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.*, **1**, 135–141.

[3] Andersson, S. A., Madigan, D., and Perlman, M. D. (1996). An alternative Markov property for chain graphs. *Proc. 12th Conf. Uncertainty in Artif. Intell.*, E. Horvitz and F. Jensen, eds. Morgan Kaufmann, San Mateo, Calif., pp. 40–48.

[4] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

[5] Bartlett, M. S. (1935). Contingency table interactions. *Suppl. J. R. Statist. Soc.*, **2**, 248–252.

[6] Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc. B*, **25**, 220–233.

[7] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.

[8] Bouckaert, R. and Studeny, M. (1995). Chain graphs: semantics and expressiveness. In *Symbolic and Qualitative Approaches to Reasoning and Uncertainty*, C. Froidevaux and J. Kohlas, eds. Springer-Verlag, Berlin, pp. 69–76.

[9] Cochran, W. G. (1965). The planning of observational studies of human populations. *J. R. Statist. Soc. A*, **128**, 234–265.

[10] Cox, D. R. (1966). Some procedures connected with the logistic qualitative response curve. In *Research Papers in Statistics: Essays in Honour of J. Neyman's 70th Birthday*, F. N. David, ed. Wiley, London, pp. 55–71.

[11] Cox, D. R. (1972). The analysis of multivariate binary data. *Appl. Statist.*, **21**, 113–120.

[12] Cox, D. R. (1997). Barlett's adjustment. In *Encyclopedia of the Statistical Sciences* Update, **1**, S. Kotz, C. B. Read and D. L. Banks, eds. Wiley, New York, pp. 43–45.

[13] Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.*, **8**, 204–218, 247–277.

[14] Cox, D. R. and Wermuth, N. (1994). Tests of linearity, multivariate normality and the adequacy of linear scores. *Appl. Statist.*, **43**, 347–355.

[15] Cox, D. R. and N. Wermuth (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation*. Chapman and Hall, London.

[16] Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov fields and log-linear models for contingency tables. *Ann. Statist.*, **8**, 522–539.

[17] Dawid, A. P. (1979). Conditional independence in statistical theory. *J. R. Statist. Soc. B*, **41**, 1–31.

[18] Dempster, A. P. (1972). Covariance selection. *Biometrics*, **28**, 157–175.

[19] Edwards, D. (1995). *Introduction to Graphical Modelling*. Springer-Verlag, New York.

[20] Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist.*, **17**, 333–353.

[21] Gibbs, W. (1902). *Elementary Principles of Statistical Mechanics*. Yale University Press, New Haven.

[22] Giesen, H., Böhmeke, W., Effler, M., Hummer, A., Jansen, R., Kötter, B., Krämer, H.-J., Rabenstein, E., and Werner, R. R. (1981). *Vom Schüler zum Studenten. Bildungslebensläufe im Längsschnitt*, Monografien zur Pädagogischen Psychologie **7**, Reinhardt, München.

[23] Goodman, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. *J. Amer. Statist. Ass.*, **65**, 226–256.

[24] Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika*, **60**, 179–192.

[25] Haberman, S. (1974). *The Analysis of Frequency Data*. University of Chicago Press. Chicago.

[26] Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, **11**, 1–12.

[27] Jöreskog, K. G. (1981). Analysis of covariance structures. *Scand. J. Statist.*, **8**, 65–92.

[28] Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Statist.*, **23**, 105–116.

[29] Koster, J. (1996). Markov properties of nonrecursive causal models. *Ann. Statist.*, **24**, 2148–2177.

[30] Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.

[31] Lauritzen, S. L., Dawid, A. P., Larsen, B., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, **20**, 491–505.

[32] Lauritzen, S. L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31–54.

[33] Leimer, H.-G. (1989). Triangulated graphs with marked vertices. In *Graph Theory in Memory of G. A. Dirac*, L. D. Andersen, et al., eds. Ann. Discrete Math. **41**, Elsevier, Amsterdam, pp. 99–123.

[34] Leimer, H.-G. (1993). Optimal decomposition by clique separators. *Discrete Math.*, **113**, 99–123.

[35] Leber, M. (1996). Die Effekte einer poststationären telefonischen Nachbetreuung auf das Befinden chronisch Schmerzkranker. Dissertation, Medical School, Universität Mainz.

[36] Markov, A. A. (1912). *Wahrscheinlichkeitsrechnung*. Teubner, Leipzig. (German translation of 2nd Russian ed., 1908).

[37] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London.

[38] Neapolitan, R. E. (1990). *Probabilistic Reasoning in Expert Systems*. Wiley, New York.

[39] Oliver, R. E. and Smith, J. Q. (1990). *Influence Diagrams, Belief Nets and Decision Analysis*. Wiley, London.

[40] Olkin, I. and Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.*, **32**, 448–465.

[41] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, Calif.

[42] Pearl, J. and Verma, T. (1988). The logic of representing dependencies by directed graphs. *Proc. 6th Conf. Amer. Ass. Artif. Intell.*, Seattle, Wash., pp. 374–379.

[43] Speed, T. P. (1979). A note on nearest-neighbour Gibbs and Markov distributions over graphs. *Sankhyā A*, **41**, 184–197.

[44] Speed, T. P. and Kiiveri, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.*, **14**, 138–150.

[45] Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G. (1993). Bayesian analysis in expert systems (with discussion). *Statist. Sci.*, **8**, 219–283.

[46] Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. *Proc. 11th Conf. Uncertainty in Artif. Intell.*, P. Besnard and S. Hanks, eds. Morgan Kaufmann, San Mateo, Calif.

[47] Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search*. Springer-Verlag, New York.

[48] Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests. *Scand. J. Statist.*, **2**, 71–79.

[49] Tarjan, R. E. (1985). Decomposition by clique separators. *Discrete Math.*, **55**, 221–232.

[50] Tukey, J. W. (1954). Causation, regression, and path analysis. In *Statistics and Mathematics in Biology*, O. Kempthorne et al., eds. Iowa State College Press, Ames, pp. 35–66.

[51] Wermuth, N. (1976). Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics*, **32**, 95–108.

[52] Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Ass.*, **75**, 963–997.

[53] Wermuth, N. (1992). On block-recursive regression equations (with discussion). *Brazil J. Probab. Statist. (Rev. Brasil. Probab. e Estatist.)*, **6**, 1–56.

[54] Wermuth, N. and Cox, D. R. (1997). On association models defined over independence graphs. *Bernoulli*, to appear.

[55] Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika*, **70**, 537–552.

[56] Wermuth, N. and Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. R. Statist. Soc. B*, **52**, 21–72.

[57] Whittaker, J. L. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, New York.

[58] Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, **9**, 60–62.

[59] Wold, H. O. (1954). Causality and econometrics, *Econometrica*, **22**, 162–177.

[60] Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism. *Genetics*, **8**, 239–255.

[61] Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.*, **5**, 161–215.

[62] Yule, G. U. (1902). Notes on the theory of association of attributes in statistics. *Biometrika*, **2**, 121–134.

(CAUSATION
CONDITIONAL INDEPENDENCE
EXPERT SYSTEMS, PROBABILISTIC
GRAPH THEORY
PATH ANALYSIS)

NANNY WERMUTH

GROUPING NORMAL MEANS

The comparison of k normal means $\mu_1, \mu_2, \ldots, \mu_k$ is typically presented in the context of analysis of variance*. The hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ is tested by an F-test* with $k - 1$ and ν degrees of freedom (df), where ν represents the df for the error mean square, s^2 say, in the ANOVA.

When the groups (treatments) whose means are to be compared are qualitative and unstructured, commonly used procedures to investigate the more specific question of which means are different are multiple comparison procedures* (MCPs), multiple range tests* (MRTs), and simultaneous testing procedures* (STPs). The basic idea of such procedures is to arrange the observed means in ascending order, say $\bar{y}_1 \leqslant \bar{y}_2 \leqslant \cdots \leqslant \bar{y}_k$, each based on n observations, by testing sequentially hypotheses of the form

$$H_{p,q} : \mu_{p+1} = \mu_{p+2} = \cdots = \mu_{p+q}$$

for $p = 0, 1, \ldots, k - q$, $q = 2, 3, \ldots, k$. This is done in at most $k - 1$ stages. At stage l, hypotheses of the form $H_{p,(k-l+1)}$ for $p = 0, 1, \ldots, l - 1$ are tested [5]. More specifically, the hypothesis $H_{p,q}$ is rejected if

$$\bar{y}_{p+q} - \bar{y}_{p+1} \geqslant c^{\alpha}_{q,\nu} s/\sqrt{n},$$

where α is the familywise error rate and $c^{\alpha}_{q,\nu}$ is an appropriate critical value.

One of the drawbacks of these procedures is that often they lead to overlapping clusters of homogeneous groups (means). This may be unsatisfactory from a practical point of view. To alleviate this problem, alternative procedures have been proposed which are embedded in STPs or make use of MCPs and clustering methods. All of these procedures are sequential and lead to nonoverlapping clusters of homogeneous groups.

Basically, the procedures can be classified in two ways: hierarchical (H) versus nonhierarchical (NH), and agglomerative (A) versus divisive (D). Here, *hierarchical* means that once a group is assigned to a cluster in one step of the procedure, then it will remain in that cluster at subsequent steps. *Agglomerative* means that at each step two clusters will be combined to form a new cluster, and *divisive* means that at each step more clusters are being formed than existed in the preceding step. Methods by several authors fall into the various categories mentioned above (see Table 1).

Table 1

	A	D
H	Calinski and Corsten (1) [1], Jolliffe [5]	Scott and Knott [6]
NH	—	Calinski and Corsten (2) [1]