CHAPTER 3:

REVIEW OF LITERATURE

The Linear Models Approach

Two major thrusts in the statistical development of measures of relations occurred during this century. A complete survey is not intended here, nor will the Chapter 1 discussion of the fundamental assumptions of linear models be repeated. Rather, NTPA will be contrasted to those major statistical developments (see Dudycha and Dudycha, 1972, for an historical perspective on behavioral statistics).

One tradition follows that of Galton and Karl Pearson. Measures arising from this tradition include the Pearson product moment coefficient and corresponding regression analysis and its extensions--i.e., factor, path, canonical analysis. This approach has been often termed, 'descriptive-correlational'. The other major tradition follows that of Sir Ronald Fisher, Jerzy Neyman and Egon Pearson, and includes analysis of variance (ANOVA), its offspring, and tests of statistical inference. This approach is often termed, 'experimental'.

While modern statisticians have shown that both approaches can be modeled by linear equations, they fundamentally differ in their aims (e.g., Kerlinger & Pedhazur, 1973). Fisher's ANOVA was in the experimental tradition. That is, one attempts by means of experimental manipulation of independent variables to identify causal factors which influence the experimental outcomes (dependent variables measured at an interval or ratio level). The method presumes random selection and assignment of subjects or cases to experimental conditions so as to minimize the possibility that spurious factors, unaccounted for in the experimental design, might influence the results in a systematic way.

In the descriptive-correlational tradition, one observes the statistical relations among variables of interest. There is no experimental control of independent variables. Fisher, a student of Karl Pearson and then a bitter rival to the end, often cited the problem of coincidence in correlational analysis that has been of concern at least since Aristotle (Analytica A Posteriori)--i.e., correlation does not imply causation. Two variables may be highly associated in a descriptive sense, but it is not safe to conclude that one causes the other. Correlation is a symmetric measure of association, and therein lies the problem. Causal relations are usually conceived as being asymmetric. That is, both:

If *A*, then *B*

and If not *A*, then not *B*

must be true in order to make a causal inference in the usual interpretation of causality. Experimentalists such as Fisher argued that the only way to tell for sure if *A* causes *B* is to manipulate *A* and observe the results. The use of an experimental and control group methodologically parallels the above two logical relationships, respectively.

The correlationist's basic reply to the experimentalists was not on logical but pragmatic grounds. Pearson and his followers claimed that it is often impractical, impossible, or sometimes unethical to manipulate certain independent variables. Hence, it is contended that the descriptive-correlational approach is a first step, although not a decisive one, to identification of potential causal factors. The present author does not take issue with this contention. However, more recently in path analysis, which is fundamentally correlational analysis, some investigators have been making causal claims from descriptive data (e.g., Asher, 1976; Blalock, 1967; 1970; Duncan, 1966; Heise, 1970; Joreskog, 1969; Wright, 1921). Causality is in effect defined to be correlation (e.g., see Asher, 1976). In other words, the causality referred to in path analysis is not the Aristotelian/Fisherian interpretation of causality.

In addition to the assumption that relations are linear or curvilinear in path analysis, it is also assumed that a theory is being tested. The theory should determine the structure of the digraph (i.e., path diagram) and sign of path coefficients prior to attempts of path analytic verification. Structural equations are defined which model the hypothesized structure of relations among factors operationalized as variables. Path coefficients are obtained using these equations and empirical data with some method of estimation such as ordinary least squares, two- or three-stage least squares, or maximum likelihood procedures. If path coefficients are statistically significant and agree in sign with those hypothesized *a priori*, then the data are said to confirm the theory under investigation. If that theory contains causal relationships among factors, these relationships are said to be confirmed. Although no experimental manipulation of factors has occurred, there is a statistically significant goodness of fit between the model represented by the structural equations and the empirical data.

While debate continues over whether causal inferences can be validly made from correlational data, the assumption is made in path analysis that relations among factors are functional in the set-theoretic sense, represented by structural equations. See Chapter 1 for the discussion of determinism in the linear models approach.

More recently, another variant of correlational analysis termed, 'time series analysis', has been developed (e.g., see Glass, Wilson, and Gottman, 1975). While time series analysis is similar in intent to NTPA, the former is nonetheless deterministic and linear, as are ANOVA, regression and path analysis.

There is also an application of ANOVA which has been termed, 'Aptitude-Treatment Interaction' (ATI) (Cronbach & Snow, 1969). In ATI what is desired are no statistically significant main effects but instead significant interactions among levels of factors. This is the converse of what is usually wanted in ANOVA. Of all the linear models approaches, ATI is probably the closest in intent to NTPA. The major differences are, however, that in NTPA no distinction is made between independent and dependent variables, none of the factors are measured at the interval or ratio level, the measurement procedures inherent in NTPA permit sequential and/or simultaneous mapping of levels of factors (i.e., categories in classifications) in time, and NTPA does not require the assumptions of linearity and additivity as does ANOVA. ATI and NTPA are similar in that an investigator seeks to verify or discover combinations of levels of factors which decrease the uncertainty of prediction of levels or values of other factors.

Contingency Analysis of Nominal and Ordinal Data

There are several variants of contingency analysis, or methods of determining relationships of variables which are measured at the nominal or ordinal level. Among those methods are chi-square (test of independence), phi, Kendall's tau-*A* and tau-*B*, lambda, gamma, Cramer's *V* and information theoretic coefficients such as *T* and *U* (e.g., Reynolds, 1977). While NTPA is similar to these methods, in that the set-theoretic concept of 'relation' is taken as the Cartesian product of two or more sets (classifications) resulting in a set of ordered pairs (or more generally, *n*-tuples), the similarity basically ends at this point. In NTPA the information theoretic concepts of classifications, each comprised of mutually exclusive and exhaustive categories, and the concepts of simultaneity and sequence of categories in time, makes NTPA more than simple cross-tabulation of nominal or ordinal level variables.

Perhaps closest in intent to NTPA is multivariate contingency analysis. The multivariate analysis of qualitative data has been of interest for some time, particularly in sociology in analysis of survey or panel data. Goodman (1970; 1972; 1973a; 1973b; 1974; 1978) has developed and presented the most comprehensive explication of this problem and some solutions to it. A general log-linear model has been proposed by Goodman (1978), which is analogous to the general linear models approach to analysis of quantitative data (e.g., see Jöreskog, 1969; 1971; 1974).

The log-linear model is predicated on the basic assumptions that observations are independent and variables are measured at a nominal or ordinal level. No assumptions are made with respect to normality of population distributions and homoscedasticity of population variances, which are required for statistical inference in quantitative analysis (ANOVA, regression, etc.), since these assumptions often cannot be met with qualitative data.

In contingency analysis the prediction of cell values from frequency data is multiplicative, not additive. An expected cell frequency is viewed in terms of parameters which estimate the so called effects of marginal distributions on that cell. For example, in a saturated model with variables A and B, with respective levels i and j, the expected cell frequency, F_{ij}, is modeled as:

$$F_{ij} = \eta\, \tau_i^A\, \tau_j^B\, \tau_{ij}^{AB}$$

where η is a constant correction factor to meet the restriction that $\Sigma_{i,j} p_{ij}=1$; and the τ values represent the so called effects of the levels of the variables and their interaction. By taking the natural logarithm of this equation, it becomes:

$$\gamma_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

where $\gamma_{ij} = \log(F_{ij})$, $\mu = \log(\eta)$, $\lambda_i^A = \log(\tau_i^A)$, etc. This equation analogously resembles that which serves as a model for ANOVA. The further assumptions are that:

$$\Sigma_i \lambda_i^A = \Sigma_j \lambda_j^B = \Sigma_i \lambda_{ij}^{AB} = \Sigma_j \lambda_{ij}^{AB} = 0$$

and

$$\{exp\ \mu\}\{\Sigma_{i,j}\ exp(\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB})\} = 1\ .$$

The saturated model, which contains all factors and their interactions, is generally of little interest by itself, since the λ (or τ) parameters can be estimated so as to fit the data perfectly. What one usually tries to do is to find an unsaturated model which is consistent with theoretical expectations and which parsimoniously provides a relatively good fit to the data. That is, in an unsaturated model some of the paramaters (factors and/or their interactions) are omitted, thus assuming they are independent (i.e., not significantly related to the prediction of cell frequencies); the remaining parameters serve to predict the cell frequencies. Goodness of fit of the model is tested by the chi square statistic with appropriate degrees of freedom.

The major problem with this approach is that more than one unsaturated model may provide a good fit to the data. Thus, an investigator must provide additional reasons for choosing one model over another--e.g., based on theoretical expectations or known temporal ordering of factors. Goodman (1978) has explicated hierarchical testing procedures to guide investigators in identification of models and has also demonstrated that the approach can be used for path analysis of qualitative variables.

While the general log-linear model is quite elegant mathematically, it nonetheless differs from NTPA in two important ways: 1) Independence of observations is assumed in constructing a contingency table in log-linear analysis. To meet this assumption cases are randomly sampled from the population of interest. The joint occurrence of the specific levels of the qualitative variables observed for a given case (i.e., unit of analysis) represents a <u>single frequency</u> in the apporpriate cell of the contingency table. Each of these cases must be independent in order to validly make statistical inferences with chi square. In NTPA independence of observations is not assumed <u>within</u> a given case (i.e., system), where repeated observations are made of the joint and sequential occurrences of events within that given system. Although a contingency table representing a set of specified nonmetric temporal paths can be constructed for each system observed, these repeated observations within each system (case) are not assumed to be independent. In fact, it is the dependencies among the occurrences of events within a system that are of interest. If systems are randomly sampled in NTPA, the proportions in the contingency tables for those systems for a set of nonmetric temporal paths are averaged using the arithmetic mean. The results in NTPA are averaged proportions with associated variances in each cell of the summary table, where there is a cell for each nonmetric temporal path in the set. 2) In NTPA joint and/or sequential occurrence of relations is explicitly considered. Thus, a variety of summary tables resembling contingency tables can be constructed for a given system, depending on how the classifications are ordered and/or combined. In other words, many different kinds of summary tables may be constructed in NTPA, depending on the temporal ordering of categories in classifications. In the general log-linear model only one complete contingency table is utilized--i.e., containing the intersection of all factors. Of course, reduced contingency tables containing a subset of the factors are derivable from the complete table.

In summary, Goodman's approach to multivariate analysis of qualitative data is similar in many respects to NTPA, but is based on assumptions not made in NTPA. The approaches are equivalent at a fundamental level only in the special case where each system is observed only once for a singular occurrence of a joint/sequential event which is characterized by categories in the classifications of interest. Then all possible single segment and single phrase multi-segment ("AND") queries are asked of the data collected on all observed systems. See Chapter 2. This would result in a complete contingency table containing the classifications for all factors, as well as derivable reduced tables, which could then be analyzed for dependence in the Goodman approach. The Goodman approach does not appear to be appropriate for repeated measures within each system, since the assumption of independence of observations is not fulfilled. If this problem could be validly solved, then the Goodman approach would provide a significant extension of NTPA. It is doubtful, however, that it can be solved, since the classical formula for conditional probability does not generally hold in NTPA for frequency measures. Functionality is not assumed in NTPA--i.e., there may be a one-to-many correspondence between elements of the domain and range of the sets of categories in classifications. See Chapter 2.

Sequential Analysis

Interest in sequential analysis or analysis of sequential processes has grown during the last half of this century. Much of the research concerning sequential analysis, outside of mathematical theory, has occurred in the fields of psychology and sociology (e.g., Coombs, Dawes & Tversky, 1970), although educational researchers have become more interested since the boom in classroom observational research that began in the early 1960's (e.g., Medley & Mitzel, 1963). The large majority of this research has utilized Markovian models.

Coombs, Dawes and Tversky (1970) have outlined the basic assumptions of Markovian modeling. A state-space approach is postulated, where the process studied is characterized as being in only one of a number of possible states at a given moment in time. A set of states is considered, $S = \{S_1, S_2, \ldots S_m\}$, at each moment in time, $\{T_1, T_2, \ldots T_n\}$. The sequence is then defined as an n-tuple of the form, $(S_i \;\&\; T_n, S_j \;\&\; T_{n+1},$

S_k & T_{n+2}, ...). Psychologists have often characterized the moments as trials (e.g., learning trials), whereas educators have been more interested in the occurrence of behaviors or point-time samples as moments (e.g., Collett & Semmel, 1973). The predominant model used has been Markovian, in which the transition between the n-1 and nth states of a sequence is considered. The Markovian model therefore is limited to two-stage chains. The major assumption for a Markovian process is that the probability of S_i & T_n, $P(S_i$ & $T_n)$, following any sequence of states is independent of all prior states in that sequence except the one that occurs at T_{n-1}. This means that the path preceding S_j & T_{n-1} is assumed to have no influence on the $P(S_j$ & T_{n-1}, S_i & $T_n)$. A Markov process is thus 'path independent'. A Markov chain is a Markov process which is independent of the trial numbers (moments in time). A Markov process is not restricted by the assumption of time independence as is a Markov chain--i.e., the location in the overall sequence observed (e.g., the beginning or end of the observation). Markov chains and processes can be conveniently characterized by a matrix of transitional probabilities, where the rows are the immediate antecedent states and the columns are the consequent states.

Collett and Semmel (1973) questioned whether Markovian assumptions were appropriate for the study of behavior sequences in educational and social environments. Sequences longer than two stages would most likely be of interest to both researchers and practitioners. It is easy to see why sequences longer than two stages have been generally avoided, since the number of possible transitions geometrically expands rapidly for longer sequences. For example, for a ten-category classification, the number of different possible five-stage sequences is 10^5, or 100,000 cells in the transition matrix. Clearly, as the length of the sequence increases or as the number of categories increases, the number of possible chains, plus shorter chains embedded within, quickly becomes unwieldy for analysis, even with modern day computer technology. Collett and Semmel (1973) sought to identify a subset of possible sequences through empirically based analysis of chains, where criteria such as focal elements, redundancy thresholds, noise thresholds and duration thresholds were utilized to reduce the magnitude of the problem.

Simultaneity of events has been largely ignored in sequential analysis. Fink (1970) asserted, for example, that sequential analysis ignores the fact that both teachers and students behave simultaneously. One solution to that problem is to create a new classification which consists of categories which characterize simultaneous combinations of student and teacher behavior (e.g., see Coombs, et al., 1970). If a teacher classification contained fifteen categories and a student classification was comprised of ten categories, then there would be 15 X 10 or 150 simultaneous (or joint) states represented by categories in the single combined classification. There are approximately 76 billion possible five-stage sequences in this 150-category classification.

In the early stages of conceptualization of NTPA (circa 1974 to 1978), the present author sought a method which could economically capture both simultaneous and sequential characteristics of events. Systematic observation (see Chapter 2) was the result of that effort. This procedure provides a very compact way of preserving both the simultaneity and sequence of events for real time observational records. Unlike the restricted, a posteriori, inductive method of Collett and Semmel (1973), NTPA requires the investigator to specify the temporal patterns of interest. Theoretical hypotheses therefore must guide inquiry about patterns in the data. Thus, length or complexity of a temporal path is not necessarily restrictive in NTPA, since the raw observational data are preserved. The analytical problem is not one of considering a very large number of cells in a transition matrix or contingency table, but one of tracing and counting only specified patterns in the observational data. The magnitude of the problem is reduced in NTPA by searching only for patterns of theoretical interest. This is in keeping with current scientific practice. A solely inductive method cannot work (Steiner, 1978).

Morgan and Messenger (1973) have developed a sequential analysis procedure (THAID) for analysis of nominal and/or ordinal level independent and dependent variables. Their procedure is somewhat analogous to multiple discriminate function analysis and multiple regression analysis with dummy coding of independent and/or dependent variables. THAID is also similar to a previous procedure, automatic interaction detection (AID) (Sonquist, Baker & Morgan, 1971), where the dependent variable is interval or ratio level and the independent variables are nominal or ordinal level.

The purpose of THAID is, through a binary splitting algorithm, to identify "... a set of subgroups characterized by a few independent variable attributes, whose dependent variable distributions are maximally different." (Morgan & Messenger, 1973, p. 2) Two statistical criteria for grouping may be used: theta and delta. Theta is a statistic which is defined as "a proportion of the sample classed correctly when using the optimal-prediction-to-the-mode strategy." (p. 12) The delta statistic is based on the criterion of "split groups whose probability distributions differ maximally from the original group and hence from each other. To account for differing split group frequencies the probability distribution 'distances' are frequency weighted." (p. 15) Morgan and Messenger (1973) claim that the functional form of the THAID model is predictive and non-additive.

Like ATI and multivariate contingency analysis, THAID is similar in intent to NTPA--i.e., to verify or discover combinations of levels of factors which reduce uncertainty of prediction of levels or values of other factors. However, THAID assumes the same level of independence of observations as does contingency analysis (see above comparison of contingency analysis and NTPA). Multiple observations of patterns within a system are not made in THAID as they are in NTPA. Most importantly, THAID uses statistical criteria (theta and delta) for identification of combinations of levels of factors which differentiate groups on the dependent variable, whereas NTPA depends on theoretical insight to determine which patterns to investigate. Clearly, THAID is limited by the number of factors and levels within factors (i.e., categories within classifications) that can be handled. The geometric expansion problem, discussed above, limits THAID but not NTPA.

Concluding Remarks

The study of sequential patterns in educational inquiry, particularly classroom observational research, is not new. For example, Flanders (1970) analyzed two-stage behavior chains in observational data obtained from his interaction analysis system, utilizing a Markovian model. Bellack, Kliebard, Hyman and Smith (1966) analyzed patterns of structuring, soliciting, responding and reacting moves of teachers and students in the classroom and found that teachers were more likely to behave in certain patterns than were students. Collett and Semmel (1973) studied the difference in transition matrices (two-stage chains) when point-time sampling was used versus real-time continuous coding and found that the former tended to yield less valid estimates of transitional probabilities. Dunkin and Biddle (1974) reviewed a number of studies in which patterns of behaviors were of concern, including teaching cycles. Medley (1979) indicated in a review of a large number of studies of teaching effectiveness that some of the variables related to student achievement were themselves patterns of teacher-student behavior. The value of sequential analysis apparently has been recognized (e.g., Dunkin & Biddle, 1974), but the general difficulty has been to develop educational theory which adequately explains and/or predicts sequences.

However, the predominate mode of analysis of relations in educational research has not been sequential analysis, but instead the linear models approach (LMA). If patterns are analyzed, the tendency is to treat a pattern itself as a variable, which in turn is related by some function to other independently measured variables (e.g., Medley, 1979). While there are notable exceptions to this trend (e.g., Bellack, et al., 1966), the majority of educational research studies tend to report relations as linear or curvilinear functions--i.e., as being deterministic.

NTPA provides a methodological framework for investigation of joint and/or sequential occurrences of events in educational systems. NTPA is grounded in set theory, probability theory and information theory. Relations are assumed to be stochastic and to exist at the level of individual systems (including individual systems-environment interactions). In NTPA relations among factors are <u>not</u> viewed as functional (i.e., deterministic) as they are in the LMA. NTPA differs from contingency analysis both in terms of temporality of factors and the level at which independence of observations is assumed. NTPA is very similar to Markovian modeling in principle. However, multiple classifications can be considered in NTPA, thus permitting the investigation of joint (simultaneous) occurrences as well as sequential occurrences of events. In addition, chains greater than length of two can be investigated through NTPA. Thus, NTPA extends Markovian theory, although the classical formula for conditional probability does not generally obtain for relative frequencies of sequences in NTPA. This may be viewed as analogous to not assuming the parallel postulate in Euclidean geometry, permitting the development of non-Euclidean geometries.

One may argue that NTPA is just another name for sequential or pattern analysis, and therefore should not be regarded as an innovation. In many respects this is true, with the exception of the methodological rigor that is added by set, probability, information and systems theory and the procedures for what has been termed, 'systematic observation' (see Chapter 2). The major issue is more fundamental. It is a matter of meta-theoretical viewpoint. If one ascribes to determinism, relations are methodologically verified by linear or curvilinear functions. Stochastic relations are ruled out. If one ascribes to a stochastic systems paradigm, however, deterministic relations are not ruled out. Rather deterministic relations are viewed as a special case of a more general stochastic systems paradigm.

In the following chapter NTPA is compared to the LMA by means of an empirical example. While the shift in viewpoint is subtle, the difference in the results from the two approaches is rather dramatic. Although the difference should be obvious from a logical viewpoint, sometimes the obvious is not easy to see. A logical comparison of NTPA and the LMA is then made in Chapter 5.