CHAPTER 1:

THE PROBLEM

Introduction to the Problem

Extant statistical models widely used in educational research and the social sciences in general are primarily linear and deterministic (Maccia and Maccia, 1976; Steiner, 1978). Analysis of variance (ANOVA), regression analysis, and their extensions (i.e., discriminant, factor, canonical and path analysis) are predicated upon deterministic relations. While statisticians rightly warn investigators that such linear models should be avoided for statistical verification of research hypotheses when the assumptions of those models are inappropriate or remain unfilled, such advice is often ignored (e.g., see Coombs, 1960; Kaplan, 1964).

Embedded within the linear models approach is a theory of measurement, where variables are independently assessed, and their relations are estimated by means of the research design and statistical procedures. This appears to follow the paradigm of Newtonian mechanics and the atomism and determinism implicit in the philosophy of Descartes and others.

It will be argued that such a world view has inherent limitations which may have obfuscated empirical results and impaired the growth of educational science. Proposed as an alternative to the deterministic linear models approach is a stochastic, systems analytic procedure developed by the author and termed herein, 'nonmetric temporal path analysis' (NTPA). Fundamentally, the difference between the two approaches is this: In the linear models approach a relation is taken as a measure of the association between two or more variables which are independently measured. In NTPA, a relation is viewed as a nonmetric temporal path. The two procedures differ at the measurement level. Put simply, a relation between theoretical concepts is verified in the linear models approach by estimating parameters of an equation which minimize errors of prediction; whereas in NTPA it is verified by means of counting occurrences of the observed relation, resulting in a relative frequency which estimates the probability of that relation. NTPA can be employed within the confines of extant statistical inference, so that is not an area of difference. What is a central issue is the manner in which conceptual or theoretical relations are empirically measured and verified.

The Linear Models Approach

Both regression analysis and ANOVA are predicated on the basic form a linear, mathematical equation (e.g., see Hays, 1973; Kerlinger and Pedhazur, 1973; Kirk, 1968):

$$Y = B + AX + E \qquad [1]$$

Variable Y is usually termed the dependent variable and variable X the independent variable. In such a view the dependent variable is assumed to be functionally related to the independent variable, in the mathematical set-theoretic sense of function. That is, each X value is assumed to determine only one predicted Y value, where A is the magnitude of the effect of X on Y. B is viewed as a constant term in the equation (the Y axis intercept in regression analysis; the pivot group mean in ANOVA). That X may be nominal level in ANOVA or regression analysis is irrelevant to the assumption of determinism. In [1] E is taken as error of measurement or residual, and it is assumed that errors are normally and randomly distributed when making statistical inferences.

Addition of exponential or interaction terms to the basic linear equation does not change the fact that the relation is still assumed to be deterministic in this approach. While such a relation may be modeled as curvilinear, it is not non-linear, as sometimes mistakenly so called. This is a linear models approach (LMA) to the measurment and verification of relations, since it fundamentally relies on the modeling of a relation by a line surface in n-dimensional space. The heart of the issue therefore seems to be linearity or determinism. We owe to Descartes the idea of the Cartesian coordinate system, as illustrated in Figure 1.

It is true that regression may be considered to be unrestricted, and hence non-linear, but that does not imply indeterminism. In the case of ANOVA where the independent variable is nominal or categorical (or for that matter interval or ratio level) determinism is still assumed. Kerlinger and Pedhazur (1973) have shown, for example, that ANOVA is a special case of the more general linear models approach, where the independent variables are typically nominal level. By use of so called 'dummy variables', categorical variables may be represented in the linear equation.

Goodman (1978) has developed an analogous procedure for multivariate analysis of qualitative/categorical data, which is based on log-linear

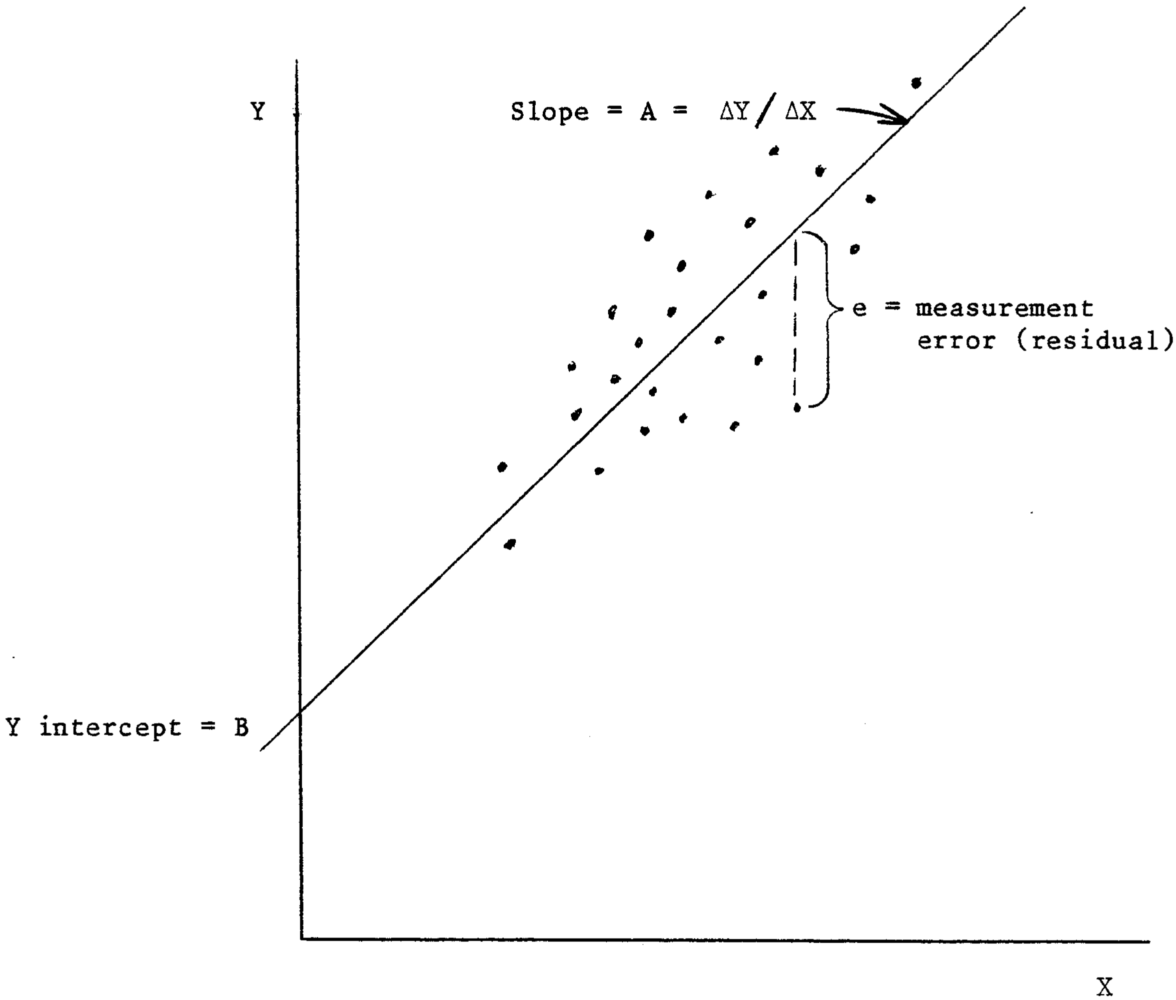

Figure 1. Graphic Illustration of Y = B + AX + E in a Cartesian Coordinate System

models. While his procedure will be reviewed in some detail in a following chapter, it differs primarily from NTPA in the level at which independence of observations is assumed. Moreover, only one type of contingency table is of concern in the log-linear models approach, whereas in NTPA a variety of types of contingency tables can be constructed for a given system, depending on how the categories are ordered or combined.

Finally, it is common in the linear models approach to estimate the strength of a relation in terms of the percent of variance of one variable or set of variables predictable from another variable or set of variables. In ANOVA eta squared or omega squared is used to estimate the strength of the statistical association between independent and dependent variables--i.e., the percent of variance of the dependent variable(s) accounted for by the independent variable(s) (Hays, 1973; Kirk, 1968). In correlational analysis, the square of the product moment coefficient is likewise an indication of the strength of the relationship between two variables in terms of the percent of variance of one variable predictable by another. In multiple correlational analysis (regression, path, canonical, discriminate and factor analysis), strength of association is similarly estimated (Kerlinger and Pedhazur, 1973). All of these approaches utilize the concept of proportional reduction of error (Costner, 1965). In essence, the variance of measurements around their statistical mean or a regression line is considered to be error of prediction. Uncertainty is reduced to the extent that knowledge of another variable or variable set reduces the error of prediction of an initial variable or variable set. Proportional reduction of error is relative to the error of prediction that would occur <u>without</u> that knowledge of values of other variables. This uncertainty in the LMA is expressed in terms of percent of variance accounted for.

Nonmetric Temporal Path Analysis (NPTA)

As an alternative to the linear models approach, the verification of a relation need not be constrained by linear and metric assumptions. In other words, a relation need not be viewed as functional in the mathematical, set-theoretic sense--i.e., modeled by an equation for a line surface. In set theory a relation is taken as the Cartesian product of two or more sets of elements. An ordered pair, or more generally *n*-tuple, symbolizes the specific joining of elements. In information theory, categories in a classification are analogous to elements in a set, with the added condition that categories are mutually exclusive and exhaustive (Atteneave, 1959; E. S. Maccia, 1963; Maccia & Maccia, 1966). Likewise, relations between categories can be modeled by the cross product of two or more classifications. By taking information as a characterization of occurrences, observed events or states of affairs can be mapped into categories in classifications (Maccia & Maccia, 1966). When time of occurrence is considered as well, such a mapping constitutes a nonmetric temporal path for each of the *n*-tuples of categories. The path is nonmetric since the mapping is not measurement in the usual sense of assigning a numerical value to a state of affairs--i.e., that the classification into which events are mapped is the set of integer or rational numbers. The path is temporal since simultaneity and sequence in time are also mapped.

When the occurrences of nonmetric temporal paths are observed and enumerated, their relative frequency can be estimated. Since probability theory can be defined in terms of set theory, and information theory can be defined in terms of both probability theory and set theory, the uncertainty of occurrences of a nonmetric temporal path can be expressed in terms of a probability measure. This procedure has been termed, 'nonmetric temporal path analysis' (NTPA) (Frick, 1980; 1982).

In summary, in the linear models approach (LMA) the strength of a relation is estimated by the linear association between two or more independently measured variables. In NTPA the strength of a relation is estimated by the relative frequency or duration of occurrences of a nonmetric temporal path. The two procedures differ at the measurement level and in the manner in which the strength of an empirical relation is derived.

The meta-theoretical difference between the LMA and NTPA concerns the assumption of determinism. Rather than arguing for determinism or indeterminism, a stochastic systems paradigm is adopted herein, and it will be shown that a deterministic system is a special case of a more general stochastic systems paradigm.

Meta-theoretical Considerations

In the history and philosophy of science, physicists faced the problem of indeterminacy in quantum mechanics, when it was discovered that the momentum and position of a sub-atomic particle could not be simultaneously determined. This brought to the forefront philosophical considerations of stochastic or indeterminate processes. More recently, these issues have been addressed in the context of general systems theory (e.g., Ackoff & Emery, 1972; von Bertalanffy, 1955; 1968; Churchman, 1968; Giere, 1973; 1976; 1979b; Maccia & Maccia, 1966; Weinberg, 1975), and in the movement away from the "Received View" or syntactic view of scientific theory construction towards a semantic or model-theoretic view (c.f., Suppe, 1973; Suppes, 1967; van Fraassen, 1980; Giere, 1979a; 1982).

Kuhn (1970) and others (e.g., Lakatos, 1968) have demonstrated historically the power of paradigms or research programs in influencing scientific thinking and resultant research methodology. What is contended here is that NTPA is more consistent with the emerging systems paradigm than with the older mechanistic view.

In the Received View, which is associated with logical positivism, events in the world are taken as inherently unconnected (c.f., Carnap, 1959; Schlick, 1959; and others from the Vienna Circle). Relations are assumed to exist in the theoretical or conceptual realm. First-order predicate calculus with identity is taken as the primitive language in the Received View, and with this theoretical terms are independently defined, and these in turn are ideally used in expression of theoretical laws or relations among concepts. Observation terms, which refer to directly observable events in the world, are linked to theoretical terms by rules of correspondence. Philosophers of science have identified many problems with the Received View, the major one being that it simply does not work in practice, but that is not of central concern here. What is noteworthy is the conception of independence of observed events. It is therefore not surprising that in the measurement theory arising in the context of this meta-theory, properties of different events are independently assessed. That is, a measure is assigned by mapping some numerical value to each event or set of events, but relations among events are not directly measured _per se_. It is in the conceptual or theoretical realm where the relations are determined--i.e., by a mathematical function indicating the relationship among independently measured properties of states of affairs.

In contradistinction to the Received View, the Semantic View of theory construction does not require first-order predicate calculus with identity nor the distinction between observational and theoretical terms and the resulting need for rules of correspondence. Any convenient extant language is taken as primitive (e.g., set theory and natural language syntax), and with this a theory is constructed as a definitional system. Quasi-isolated systems are assumed to exist in the world, and causal connections among parts of the system are to be modeled by theory. A theory is taken to model a type of system, independent of any particular interpretation or verification of the model.

Both views require empirical testing of theory for verification (observational statements in the Received View, derived hypotheses in the Semantic View), so that is not a central issue, if science is conceived as both a theoretical and empirical endeavor, as it commonly is. Problems of verification still remain in the Semantic View, but they are considered to be distinct from problems of definition. The Semantic View is preferable to the Received View since the former includes the latter. That is, the Semantic View is not restricted to the first-order predicate calculus with identity, but does not preclude it as a possible primitive language and syntax for theoretical definition.

NTPA is more consonant with the Semantic View than the Received View. In particular, the measurement theory implicit in NTPA permits direct observational measures of theoretical relations, not statistical ones. Statistics become relevant when inference or generalization is of concern in NTPA.

General systems theory has emerged as a paradigm which appears to be an alternative to the mechanistic or atomistic paradigm (c.f., von Bertalanffy, 1968; Maccia & Maccia, 1976). Giere (1973; 1976; 1979) and Weinberg (1975), for example, take a state-space view of systems. Relations are assumed to exist among states within individual systems. Giere distinguishes between deterministic and stochastic system processes. A system process is deterministic if each initial system state is associated with only one posterior state. A system process is stochastic if an initial state is associated with two or more posterior states. That is, a posterior state is not uniquely determined by a prior state in a stochastic process; rather there is a distribution of posterior states which have associated probabilities (or propensities). Therefore, it can be seen that a deterministic process is a special case where a posterior state has a probability of one, given an initial state, and the other posterior states have zero probabilities.

Maccia and Maccia (1966) further explicated general systems theory, where set theory, information theory, and digraph theory were used to define a general systems theory model, termed 'SIGGS'. It will be seen below that these theoretical concepts from SIGGS are fundamental to the definitions and derived measurement processes of NTPA.

Statement of the Problem

The linear models approach (LMA) appears to have evolved from Newtonian mechanics and the physical sciences in which deterministic assumptions and linear modeling have proved to be generally fruitful.

Verificational procedures in the social sciences, including educational research, appear to have been adopted largely from the physical science paradigm. However, educational research involves the study of human beings in a teaching-studenting process. Educational systems are organismic, not mechanistic (Steiner, 1978). Organismic systems are stochastic, not deterministic--i.e., there is uncertainty of final states with respect to any given initial state. If educational relations (i.e., system processes) are indeed stochastic, then the LMA cannot adequately verify them, since it is predicated upon deterministic assumptions. NTPA can be used to verify stochastic relations. Thus, NTPA would appear to be more adequate than the LMA for verifying stochastic relations.

The major question to be addressed herein is: Is NTPA more adequate than the LMA in the verification of stochastic educational relations? If so, the ramifications for educational research methodology are patent. If the LMA has obfuscated the development of knowledge about education through inquiry, then a more adequate methodology is likely to result in an increase in the rate of development of scientific and praxiological knowledge of education.

It will be shown that NTPA is logically inclusive of and less restrictive than the linear models approach, and therefore NTPA is more adequate than the linear models approach. In short, NTPA is not restricted by the assumption that relations are deterministic. Moreover, the fruitfulness of NTPA will be demonstrated through use of an empirical example from educational research and compared to that of the LMA in terms of conclusions possible from the data.

Overview of Following Chapters

In Chapter 2 NTPA is presented first by example. Then NTPA is formally defined. It is necessary to explicate NTPA before proceeding further, since NTPA is considered to be a new methodology.

In Chapter 3 extant methods of measuring relations are reviewed, including the LMA and contingency analysis of nominal level variables. Sequential analysis is then reviewed, particularly analysis of Markovian chains and processes. Each of these methods is contrasted to NTPA for similarities and differences.

In Chapter 4 the design and methods of gathering empirical data to compare NTPA and the LMA are explicated. The empirical results from both approaches are presented and compared.

In Chapter 5 the differences in the results and the two approaches are discussed in terms of meta-theoretical adequacy and fruitfulness.