Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

INFERENCE IN LINEAR MODELS

BY T. W. ANDERSON*
Stanford University

A general linear model can be written as $Y = XB' + U$, where Y is an $N \times p$ matrix of observable dependent variables, X is an $N \times q$ matrix of independent variables, B' is a $q \times p$ matrix of parameters, and U is an $N \times p$ matrix of unobservable random variables. The elements of X may be observable or alternatively unobservable (that is, latent); they may be nonstochastic or stochastic. The model includes regression, linear functional and structural relations, multivariate analysis of variance, factor analysis, and some simultaneous equations models. This paper considers the relationships between various models and presents methods of estimating the parameters under various conditions. Testing hypotheses about the rank of XB' (the dimensionality of the latent variables when X is not observed) are also treated.

1. A Linear Model. In this paper we consider a general linear model in multivariate analysis that includes regression models, multivariate analysis of variance (MANOVA) models, and factor analysis models. Some of these models go by names of linear functional relationships, linear structural relationships, and canonical correlations. An attempt will be made to use a unified approach to these models.

Suppose we observe the $p \times 1$ vectors $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$. A linear model is given by

$$\boldsymbol{y}_\alpha = \boldsymbol{B}\boldsymbol{x}_\alpha + \boldsymbol{u}_\alpha, \qquad \alpha = 1, \cdots, N, \tag{1.1}$$

where $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N$ are unobservable random $p \times 1$ vectors; we suppose them to be independently identically distributed (iid) with

$$\mathcal{E}\boldsymbol{u}_\alpha = \boldsymbol{0}, \qquad \mathcal{E}\boldsymbol{u}_\alpha\boldsymbol{u}_\alpha' = \boldsymbol{\Sigma}_u. \tag{1.2}$$

The $p \times q$ matrix $\boldsymbol{B}$ consists of parameters, and $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ are $q \times 1$ vectors. We shall consider these vectors as observed or alternatively as unobserved (or latent); they may be fixed (that is, nonstochastic) or alternatively random (or stochastic). The model can also be written

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}' + \boldsymbol{U}, \tag{1.3}$$

---

* This paper was prepared under National Science Foundation Grant DMS 89-04851 at Stanford University.

AMS 1980 Subject Classifications: 62H15, 62H25, 62J05, 62J10.

Key words and phrases: Factor analysis, latent variables, linear functional, MANOVA, rank, regression and structural relations.