Springer Texts in Statistics

Ronald Christensen

Linear Models for Multivariate, Time Series, and Spatial Data

Springer-Verlag
世界图书出版公司

Ronald Christensen

Linear Models for Multivariate, Time Series, and Spatial Data

With 40 Illustrations

Springer-Verlag
New York Berlin Heidelberg London
Paris Tokyo Hong Kong Barcelona
世界图书出版公司
北京・广州・上海・西安

Ronald Christensen
Department of Mathematics and Statistics
University of New Mexico
Albuquerque, NM 87131
USA

Mathematics Subject Classification: 62H17

Library of Congress Cataloging-in-Publication Data
Christensen, Ronald, 1951-
Linear models for multivariate, time series, and spatial data/
Ronald Christensen.
p. cm.
Includes bibliographical references and index.
ISBN 0-387-97413-X
1. Linear models. (Statistics) I. Title.
QA279.C477 1990 90-10377
519.5-dc20 CIP

Reprinted by World Publishing Corporation, Beijing, 1993
for distribution and sale in The People's Republic of China only
ISBN 7-5062-1433-4

ISBN 0-387-97413-X Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-97413-X Springer-Verlag Berlin Heidelberg New York

Preface

This is a companion volume to *Plane Answers to Complex Questions: The Theory of Linear Models*. It consists of six additional chapters written in the same spirit as the last six chapters of the earlier book. Brief introductions are given to topics related to linear model theory. No attempt is made to give a comprehensive treatment of the topics. Such an effort would be futile. Each chapter is on a topic so broad that an in depth discussion would require a book-length treatment.

People need to impose structure on the world in order to understand it. There is a limit to the number of unrelated facts that anyone can remember. If ideas can be put within a broad, sophisticatedly simple structure, not only are they easier to remember but often new insights become available. In fact, sophisticatedly simple models of the world may be the only ones that work. I have often heard Arnold Zellner say that, to the best of his knowledge, this is true in econometrics. The process of modeling is fundamental to understanding the world.

In statistics, the most widely used models revolve around linear structures. Often the linear structure is exploited in ways that are peculiar to the subject matter. Certainly this is true of frequency domain time series and geostatistics. The purpose of this volume is to take three fundamental ideas from standard linear model theory and exploit their properties in examining multivariate, time series, and spatial data. In decreasing order of importance to the presentation, the three ideas are: best linear prediction, projections, and Mahalanobis's distance. (Actually, Mahalanobis's distance is a fundamentally multivariate idea that has been appropriated for use in linear models.) Numerous references to results in *Plane Answers* are made. Nevertheless, I have tried to make this book as independent as possible. Typically, when a result from *Plane Answers* is needed not only is the reference given but also the result itself. Of course, for proofs of these results the reader will have to refer to the original source.

I want to reemphasize that this is a book about linear models. It is not traditional multivariate analysis, time series, or geostatistics. Multivariate linear models are viewed as linear models with a nondiagonal covariance matrix. Discriminant analysis is related to the Mahalanobis distance and multivariate analysis of variance. Principal components are best linear predictors. Frequency domain time series involves linear models with a peculiar design matrix. Time domain analysis involves models that are linear in the parameters but have random design matrices. Best linear predictors are used for forecasting time series; they are also fundamental to the estimation techniques used in time domain analysis. Spatial data analysis involves linear models in which the covariance matrix is modeled from the data; a primary objective in analyzing spatial data is making best linear unbiased predictions of future observables. While other approaches to these problems may yield different insights, there is value in having a unified approach to looking at these problems. Developing such a unified approach is the purpose of this book.

There are two well-known models with linear structure that are conspicuous by their absence in my two volumes on linear models. One is Cox's (1972) proportional hazards model. The other is the generalized linear model of Nelder and Wedderburn (1972). The proportional hazards methodology is a fundamentally nonparametric technique for dealing with censored data having linear structure. The emphasis on nonparametrics and censored data would make its inclusion here awkward. The interested reader can see Kalbfleisch and Prentice (1980). Generalized linear models allow the extension of linear model ideas to many situations that involve independent nonnormally distributed observations. Beyond the presentation of basic linear model theory, these volumes focus on methods for analyzing correlated observations. While it is true that generalized linear models can be used for some types of correlated data, such applications do not flow from the essential theory. McCullagh and Nelder (1989) give a detailed exposition of generalized linear models and Christensen (1990) contains a short introduction.

ACKNOWLEDGMENTS

I would like to thank MINITAB for providing me with a copy of release 6.1.1, BMDP for providing me with copies of their programs 4M, 1T, 2T, and 4V, and Dick Lund for providing me with a copy of MSUSTAT. Nearly all of the computations were performed with one of these programs. Many were performed with more than one.

I would not have tackled this project but for Larry Blackwood and Bob Shumway. Together Larry and I reconfirmed, in my mind anyway, that multivariate analysis is just the same old stuff. Bob's book put an end to a specter that has long haunted me: a career full of half-hearted attempts at figuring out basic time series analysis.

At my request, Ed Bedrick, Bert Koopmans, Wes Johnson, Bob Shumway, and Dale Zimmerman tried to turn me from the errors of my ways. I sincerely thank them for their valuable efforts. The reader must judge how successful they were with a recalcitrant subject. As always, I must thank my editors Steve Fienberg and Ingram Olkin for their suggestions. Jackie Damrau did an exceptional job in typing the first draft of the manuscript.

Finally, I have to recognize the contribution of Magic Johnson. I was

so upset when the 1987-88 Lakers won a second consecutive NBA title that I began writing this book in order to block the mental anguish. I am reminded of Woody Allen's dilemma: is the importance of life more accurately reflected in watching *The Sorrow and the Pity* or in watching the Knicks? (In my case, the Jazz and the Celtics.) It's a tough call. Perhaps life is about actually making movies and doing statistics.

Ronald Christensen
Albuquerque, New Mexico
April 19, 1990

BMDP Statistical Software is located at 1440 Sepulveda Boulevard, Los Angeles, CA 90025, telephone: (213) 479-7799.

MINITAB is a registered trademark of Minitab, Inc., 3081 Enterprise Drive, State College, PA 16801, telephone: (814) 238-3280, telex: 881612.

MSUSTAT is marketed by the Research and Development Institute Inc., Montana State University, Bozeman, MT 59717-0002, Attn: R.E. Lund.

Contents

Chapter I

Multivariate Linear Models

Chapters I, II, and III examine topics in multivariate analysis. Specifically, they discuss multivariate linear models, discriminant analysis, principal components, and factor analysis. The basic ideas behind these subjects are closely related to linear model theory. Multivariate linear models are simply linear models with more than one dependent variable. Discriminant analysis is closely related to both Mahalanobis's distance (cf. Christensen, 1987, Section XIII.1) and multivariate one-way analysis of variance. Principal components are user-constructed variables which are best linear predictors (cf. Christensen, 1987, Section VI.3) of the original data. Factor analysis has ties to both multivariate linear models and principal components.

These three chapters are introductory in nature. The discussions benefit from the advantage of being based on linear model theory. They suffer from the disadvantage of being relatively brief. More detailed discussions are available in numerous other sources, e.g., Anderson (1984), Arnold (1981), Dillon and Goldstein (1984), Eaton (1983), Gnanadesikan (1977), Johnson and Wichern (1988), Mardia, Kent, and Bibby (1979), Morrison (1976), Muirhead (1982), Press (1982), and Seber (1984).

As mentioned above, the distinction between multivariate linear models and standard (univariate) linear models is simply that multivariate linear models involve more than one dependent variable. Let the dependent variables be $y_1, \ldots, y_q$. If n observations are taken on each dependent variable, we have $y_{i1}, \ldots, y_{iq}$, $i = 1, \ldots, n$. Let $Y_1 = [y_{11}, \ldots, y_{n1}]'$ and, in general, $Y_h = [y_{1h}, \ldots, y_{nh}]'$, $h = 1, \ldots, q$. For each h, the vector Y_h is the vector of n responses on the variable y_h and can be used as the response vector for a linear model. For $h = 1, \ldots, q$, write the linear model

$$Y_h = X\beta_h + e_h\ ,\ \mathrm{E}(e_h) = 0\ ,\ \mathrm{Cov}(e_h) = \sigma_{hh} I \tag{1}$$

where X is a known $n \times p$ matrix that is the same for all dependent variables, but β_h and the error vector $e_h = [e_{1h}, \ldots, e_{nh}]'$ are peculiar to the dependent variable.

The multivariate linear model consists of fitting the q linear models simultaneously. Write the matrices

$$\begin{aligned} Y_{n\times q} &= [Y_1, \ldots, Y_q]\ , \\ B_{p\times q} &= [\beta_1, \ldots, \beta_q] \end{aligned}$$

and

$$e_{n\times q} = [e_1, \ldots, e_q]\ .$$

The multivariate linear model is

$$Y = XB + e. \tag{2}$$

The key to the analysis of the multivariate linear model is the random nature of the $n \times q$ error matrix $e = [e_{ih}]$. At a minimum, we assume that $\mathrm{E}(e) = 0$ and

$$\mathrm{Cov}(e_{ih}, e_{i'h'}) = \begin{cases} \sigma_{hh'} & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases}.$$

Let

$$\delta_{ii'} = \begin{cases} 1 & \text{if } i = i' \\ 0 & \text{if } i \neq i', \end{cases}$$

then the covariances can be written simply as

$$\mathrm{Cov}(e_{ih}, e_{i'h'}) = \sigma_{hh'}\delta_{ii'}.$$

To construct tests and confidence regions we assume that the e_{ij}'s have a multivariate normal distribution with the previously indicated mean and covariances. Note that this covariance structure implies that the error vector in model (1) has $\mathrm{Cov}(e_h) = \sigma_{hh}I$ as indicated previously.

An alternative but equivalent way to state the multivariate linear model is by examining the rows of model (2). Write

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix},$$

$$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix},$$

and

$$e = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_n' \end{bmatrix}.$$

The multivariate linear model is

$$y_i' = x_i'B + \varepsilon_i'$$

$i = 1, \ldots, n$. The error vector ε_i has the properties

$$\mathrm{E}(\varepsilon_i) = 0,$$

$$\mathrm{Cov}(\varepsilon_i) = \Sigma_{q \times q} = [\sigma_{hh'}],$$

and for $i \neq j$

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0.$$

To construct tests and confidence regions, the vectors ε_i are assumed to have independent multivariate normal distributions.

EXERCISE 1.1. For any two columns of Y say Y_r and Y_s, show that $\mathrm{Cov}(Y_r, Y_s) = \sigma_{rs}^2 I$.

I.1 Estimation

The key to estimation in the multivariate linear model is rewriting the model as a univariate linear model. The model

$$Y = XB + e \;,\; \mathrm{E}(e) = 0 \;,\; \mathrm{Cov}(e_{ih}, e_{i'h'}) = \sigma_{hh'}\delta_{ii'} \tag{1}$$

can be rewritten as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix} = \begin{bmatrix} X & 0 & \cdots & 0 \\ 0 & X & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & & X \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_q \end{bmatrix} \tag{2}$$

where the error vector has mean zero and covariance matrix

$$\begin{bmatrix} \sigma_{11}I_n & \sigma_{12}I_n & \cdots & \sigma_{1q}I_n \\ \sigma_{12}I_n & \sigma_{22}I_n & \cdots & \sigma_{2q}I_n \\ \vdots & \vdots & & \vdots \\ \sigma_{1q}I_n & \sigma_{2q}I_n & \cdots & \sigma_{qq}I_n \end{bmatrix}. \tag{3}$$

Recalling that the Vec operator (cf. Christensen, 1987, Definition B.6) stacks the columns of a matrix, the dependent variable in model (2) is precisely $\mathrm{Vec}(Y)$. Similarly, the parameter vector and the error vector are $\mathrm{Vec}(B)$ and $\mathrm{Vec}(e)$. The design matrix in (2) can be rewritten using Kronecker products (cf. Christensen, 1987, Definition B.5). The design matrix is $I_q \otimes X$ where I_q is a $q \times q$ identity matrix. Model (2) can now be rewritten as

$$\mathrm{Vec}(Y) = [I_q \otimes X]\mathrm{Vec}(B) + \mathrm{Vec}(e). \tag{4}$$

The first two moments of $\mathrm{Vec}(e)$ are

$$\mathrm{E}[\mathrm{Vec}(e)] = 0$$

and, rewriting (3),

$$\mathrm{Cov}[\mathrm{Vec}(e)] = \Sigma \otimes I_n. \tag{5}$$

EXERCISE 1.2. Show that $[A \otimes B][C \otimes D] = [AC \otimes BD]$ where the matrices are of conformable sizes.

For estimation, the nice thing about model (1) is that least squares estimates are optimal. In particular, it will be shown that optimal estimation is based on

$$\hat{Y} = X\hat{B} = MY$$

where $M = X(X'X)^{-}X'$ is, as always, the perpendicular projection operator onto the column space of X, $C(X)$. This is a simple generalization of the univariate linear model results of Christensen (1987, Chapter II). To show that least squares estimates are best linear unbiased estimates, (BLUE's), apply Christensen's (1987) Theorem 10.4.5 to model (2). Theorem 10.4.5 states that for a univariate linear model $Y_{n\times 1} = X\beta + e, \mathrm{E}(e) = 0, \mathrm{Cov}(e) = \sigma^2 V$, least squares estimates are BLUE's if and only if $C(VX) \subset C(X)$.

The design matrix in (2) is $[I_q \otimes X]$. The covariance matrix is $[\Sigma \otimes I_n]$. We need to show that $C([\Sigma \otimes I_n][I_q \otimes X]) \subset C([I_q \otimes X])$. Using either Exercise 1.2 or simply using the forms given in (2) and (3)

$$\begin{aligned} [\Sigma \otimes I_n][I_q \otimes X] &= [\Sigma \otimes X] \\ &= \begin{bmatrix} \sigma_{11}X & \cdots & \sigma_{1q}X \\ \vdots & & \vdots \\ \sigma_{1q}X & \cdots & \sigma_{qq}X \end{bmatrix} \\ &= [I_q \otimes X][\Sigma \otimes I_p]. \end{aligned}$$

Recalling that $C(RS) \subset C(R)$ for any conformable matrices R and S, it is clear that

$$C([\Sigma \otimes I_n][I_q \otimes X]) = C([I_q \otimes X][\Sigma \otimes I_p]) \subset C([I_q \otimes X]).$$

Applying Christensen's Theorem 10.4.5 establishes that least squares estimates are best linear unbiased estimates.

To find least squares estimates, we need the perpendicular projection operator onto $C([I_q \otimes X])$. The projection operator is

$$P = [I_q \otimes X]\left([I_q \otimes X]'[I_q \otimes X]\right)^{-}[I_q \otimes X].$$

Because $[A \otimes B]' = [A' \otimes B']$, we have

$$\begin{aligned} [I_q \otimes X]'[I_q \otimes X] &= [I_q \otimes X'][I_q \otimes X] \\ &= [I_q \otimes X'X]. \end{aligned}$$

It is easily seen from the definition of a generalized inverse that

$$[I_q \otimes X'X]^{-} = [I_q \otimes (X'X)^{-}].$$

It follows that

$$\begin{aligned} P &= [I_q \otimes X][I_q \otimes (X'X)^-][I_q \otimes X'] \\ &= [I_q \otimes X(X'X)^- X'] \\ &= [I_q \otimes M]. \end{aligned}$$

By Christensen (1987, Theorem 2.2.1), in a univariate linear model $Y_{n\times 1} = X\beta + e$ least squares estimates $\hat{\beta}$ satisfy $X\hat{\beta} = MY_{n\times 1}$; thus for the univariate linear model (4), least squares estimates of $\mathrm{Vec}(B)$, say $\mathrm{Vec}(\hat{B})$, satisfy

$$[I_q \otimes X]\mathrm{Vec}(\hat{B}) = [I_q \otimes M]\mathrm{Vec}(Y)\,,$$

i.e.,

$$\begin{bmatrix} X\hat{\beta}_1 \\ \vdots \\ X\hat{\beta}_q \end{bmatrix} = \begin{bmatrix} MY_1 \\ \vdots \\ MY_q \end{bmatrix}.$$

In terms of the multivariate linear model (1), this is equivalent to

$$X\hat{B} = MY.$$

Maximum Likelihood Estimates

Write the matrices Y and X using their component rows,

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}.$$

To find maximum likelihood estimates (MLE's), we assume that Σ is nonsingular. We also assume that the rows of Y are independent and $y_i \sim N(B'x_i, \Sigma)$. The likelihood function for Y is

$$L(XB, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-q/2} |\Sigma|^{-1/2} \exp\left[-(y_i - B'x_i)'\Sigma^{-1}(y_i - B'x_i)/2\right]$$

and the log of the likelihood function is

$$\ell(XB, \Sigma) = -\frac{nq}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(y_i - B'x_i)'\Sigma^{-1}(y_i - B'x_i).$$

Consider model (2). As for any other univariate linear model, if the nonsingular covariance matrix is fixed, then the MLE of $[I_q \otimes X]\mathrm{Vec}(B)$ is the same as the BLUE. As we have just seen, least squares estimates are BLUE's. The least squares estimate of XB does not depend on the covariance matrix; hence, for any value of Σ, $X\hat{B} = MY$ maximizes the likelihood function. It remains only to find the MLE of Σ.

The log-likelihood, and thus the likelihood, are maximized for any Σ by substituting a least squares estimate for B. Write $\hat{B} = (X'X)^{-}X'Y$. We need to maximize

$$\ell(X\hat{B},\Sigma) = -\frac{nq}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(y_i - Y'X(X'X)^{-}x_i)'\Sigma^{-1}(y_i - Y'X(X'X)^{-}x_i)$$

subject to the constraint that Σ is positive definite. The last term on the right-hand side can be simplified. Define the $n \times 1$ vector

$$\rho_i = (0,\ldots,0,1,0,\ldots,0)'$$

with the 1 in the ith place.

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - Y'X(X'X)^{-}x_i)'\Sigma^{-1}(y_i - Y'X(X'X)^{-}x_i)\\
&= \sum_{i=1}^{n}\rho_i'(Y - X(X'X)^{-}X'Y)\Sigma^{-1}(Y' - Y'X(X'X)X')\rho_i\\
&= \sum_{i=1}^{n}\rho_i'(I-M)Y\Sigma^{-1}Y'(I-M)\rho_i\\
&= \mathrm{tr}[(I-M)Y\Sigma^{-1}Y'(I-M)]\\
&= \mathrm{tr}[\Sigma^{-1}Y'(I-M)Y].
\end{aligned}$$

Thus our problem is to maximize

$$\ell(X\hat{B},\Sigma) = -\frac{nq}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\mathrm{tr}[\Sigma^{-1}Y'(I-M)Y]. \tag{6}$$

We will find the maximizing value by setting all the partial derivatives (with respect to the σ_{ij}'s) equal to zero. To find the partial derivatives, we need part (3) of Proposition 12.4.1 in Christensen (1987) and a variation on part (4) of the proposition, cf. Exercise 1.8.14. The variation on part (4) is that

$$\begin{aligned}
\frac{\partial}{\partial\sigma_{ij}}\log|\Sigma| &= \mathrm{tr}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\sigma_{ij}}\right] \qquad (7)\\
&= \mathrm{tr}\left[\Sigma^{-1}T_{ij}\right].
\end{aligned}$$

where the symmetric $q \times q$ matrix T_{ij} has ones in row i column j and row j column i and zeros elsewhere. Part (3) of Christensen's Proposition 12.4.1

gives

$$\frac{\partial}{\partial \sigma_{ij}} \Sigma^{-1} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \sigma_{ij}} \Sigma^{-1} = -\Sigma^{-1} T_{ij} \Sigma^{-1} . \tag{8}$$

We need one final result involving the derivative of a trace. Let $A(s) = [a_{ij}(s)]$ be an $r \times r$ matrix function of the scalar s.

$$\frac{d}{ds}\text{tr}[A(s)] = \frac{d}{ds}[a_{11}(s) + \cdots + a_{rr}(s)] = \sum_{i=1}^{r} \frac{da_{ii}(s)}{ds} = \text{tr}\left[\frac{dA(s)}{ds}\right] . \tag{9}$$

From (8), (9), and the chain rule

$$\frac{\partial}{\partial \sigma_{ij}}\text{tr}[\Sigma^{-1} Y'(I-M)Y] = \text{tr}\left[\frac{\partial}{\partial \sigma_{ij}}\{\Sigma^{-1} Y'(I-M)Y\}\right] = \text{tr}\left[\left\{\frac{\partial \Sigma^{-1}}{\partial \sigma_{ij}}\right\} Y'(I-M)Y\right] = \text{tr}[-\Sigma^{-1} T_{ij} \Sigma^{-1} Y'(I-M)Y] . \tag{10}$$

Applying (7) and (10) to (6), we get

$$\frac{\partial}{\partial \sigma_{ij}} \ell(X\hat{B}, \Sigma) = -\frac{n}{2}\text{tr}[\Sigma^{-1} T_{ij}] + \frac{1}{2}\text{tr}[\Sigma^{-1} T_{ij} \Sigma^{-1} Y'(I-M)Y] .$$

Setting the partial derivatives equal to zero leads to finding a positive definite matrix Σ that solves

$$\text{tr}[\Sigma^{-1} T_{ij}] = \text{tr}[\Sigma^{-1} T_{ij} \Sigma^{-1} Y'(I-M)Y/n] \tag{11}$$

for all i and j.

Let $\hat{\Sigma} = \frac{1}{n} Y'(I-M)Y$; this is clearly nonnegative definite (positive semidefinite). If $\hat{\Sigma}$ is positive definite, then $\hat{\Sigma}$ is our solution. Substituting $\hat{\Sigma}$ for Σ in (11) gives

$$\text{tr}[\hat{\Sigma}^{-1} T_{ij}] = \text{tr}[\hat{\Sigma}^{-1} T_{ij} \hat{\Sigma}^{-1} Y'(I-M)Y/n] = \text{tr}[\hat{\Sigma}^{-1} T_{ij}] .$$

Obviously this holds for all i and j. Moreover, under weak conditions $\hat{\Sigma}$ is positive definite with probability one. (See the discussion following Theorem 1.2.2.)

Unbiased Estimation of Σ

The MLE $\hat{\Sigma}$ is a biased estimate just as the MLE of the variance in a standard univariate linear model is biased. (Note that the univariate linear model is just the special case where $q = 1$.) The usual unbiased estimate of Σ does not depend on the assumption of normality and is generalized from the univariate result. An unbiased estimate of Σ is

$$S = Y'(I - M)Y/[n - r(X)].$$

To see this, consider the i, j element of $Y'(I - M)Y$.

$$\begin{aligned}
\mathrm{E}[Y_i'(I-M)Y_j] &= \mathrm{E}[(Y_i - X\beta_i)'(I-M)(Y_j - X\beta_j)] \\
&= \mathrm{E}\{\mathrm{tr}[(Y_i - X\beta_i)'(I-M)(Y_j - X\beta_j)]\} \\
&= \mathrm{E}\{\mathrm{tr}[(I-M)(Y_j - X\beta_j)(Y_i - X\beta_i)']\} \\
&= \mathrm{tr}\{\mathrm{E}[(I-M)(Y_j - X\beta_j)(Y_i - X\beta_i)']\} \\
&= \mathrm{tr}\{(I-M)\mathrm{E}[(Y_j - X\beta_j)(Y_i - X\beta_i)']\} \\
&= \mathrm{tr}\{(I-M)\mathrm{Cov}(Y_j, Y_i)\} \\
&= \mathrm{tr}\{(I-M)\sigma_{ji}I\} \\
&= \sigma_{ij}(n - r(X)).
\end{aligned}$$

Thus, each element of S in an unbiased estimate of the corresponding element of Σ.

Example 1.1.1. *Partial Correlation Coefficients*

Partial correlations were discussed in Christensen (1987, Section VI.5). Suppose we have n observations on two dependent variables y_1, y_2 and $p-1$ independent variables $x_1, \ldots, x_{p-1}$. Write

$$Y = \begin{bmatrix} y_{11} & y_{12} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{bmatrix} = [Y_1, Y_2]$$

and

$$Z = \begin{bmatrix} x_{11} & \cdots & x_{1\,p-1} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{n\,p-1} \end{bmatrix}.$$

Write a multivariate linear model as

$$Y = [J, Z]B + e$$

where J is an $n \times 1$ vector of 1's. As discussed above, the unbiased estimate of Σ is $S = [s_{ij}]$ where

$$\begin{aligned}
S &= Y'(I-M)Y/[n - r(X)] \\
&= \frac{1}{n - r(X)} \begin{bmatrix} Y_1'(I-M)Y_1 & Y_1'(I-M)Y_2 \\ Y_2'(I-M)Y_1 & Y_2'(I-M)Y_2 \end{bmatrix}.
\end{aligned}$$

From Christensen (1987, Section VI.5), the sample partial correlation coefficient is

$$\begin{aligned} r_{y\cdot x} &= \frac{Y_1'(I-M)Y_2}{[Y_1'(I-M)Y_1\, Y_2'(I-M)Y_2]^{1/2}} \\ &= \frac{s_{12}}{\sqrt{s_{11}s_{22}}}. \end{aligned}$$

The sample partial correlation coefficient is just the sample correlation coefficient as estimated in a multivariate linear model in which the effects of the x variables have been eliminated.

I.2 Testing Hypotheses

Consider testing the multivariate model

$$Y = XB + e \tag{1}$$

against a reduced model

$$Y = X_0\Gamma + e \tag{2}$$

where $C(X_0) \subset C(X)$ and the elements of e are multivariate normal. The covariance matrix $[\Sigma \otimes I_n]$ from model (1.1.2) is unknown, so standard univariate methods of testing do not apply. Let $M_0 = X_0(X_0'X_0)^-X_0'$ be the perpendicular projection operator onto $C(X_0)$. Multivariate tests of model (2) versus model (1) are based on the hypothesis statistic

$$H \equiv Y'(M - M_0)Y$$

and the error statistic

$$E \equiv Y'(I - M)Y.$$

These statistics look identical to the sums of squares used in univariate linear models. The difference is that the univariate sums of squares are scalars, while in multivariate models these statistics are matrices. The matrices have diagonals that consist of sums of squares for the various dependent variables and off-diagonals that are sums of cross-products of the different dependent variables.

For univariate models, the test statistic is proportional to the scalar $Y'(M - M_0)Y[Y'(I-M)Y]^{-1}$. For multivariate models, the test statistic is often taken as a function of the matrix $Y'(M - M_0)Y[Y'(I-M)Y]^{-1}$ or some closely related matrix. For multivariate models, there is no one test statistic that is the universal standard for performing tests. Various test statistics are discussed in the next subsection.