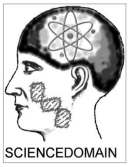
British Journal of Mathematics & Computer Science
11(6): 1-18, 2015, Article no.BJMCS.20493
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org



Relationships between Linear Statistical Models

Vladimir G. Panov[1*] **and Anatoly N. Varaksin**[1]

[1] *Institute of Industrial Ecology of Ural Branch of RAS, S. Kovalevskaya, 20, Ekaterinburg, 620990, Russian Federation.*

Article Information

DOI: 10.9734/BJMCS/2015/20493
Editor(s):
(1) Vyacheslav Pivovarchik, Department of Applied Mathematics and Computer Science, South-Ukrainian National Pedagogical University, Ukraine.
Reviewers:
(1) S. E. Fadugba, Ekiti State University, Ado Ekiti, Nigeria
(2) Kumar Dookhitram, University of Technology, Mauritius.
Complete Peer review History: http://sciencedomain.org/review-history/11390

Original Research Article

Received: 30 July 2015
Accepted: 02 September 2015
Published: 15 September 2015

Abstract

To study the relationship between the linear statistical models we used methods of linear algebra, Hilbert spaces and statistics. It was found that there is a linear relationship between linear statistical models which is expressed by a matrix equality. Several corollaries are derived and discussed, and a new interpretation is proposed for the parameters of linear statistical model. The given relation between the linear statistical models may be useful for both theoretical analysis of statistical models and interpretation of applied statistical models, in particular, to analyze the impact of confounders.

Keywords: Linear statistical model; dependency; matrix equality; interpretation of linear model parameters; Cochran's multivariate formula.

2010 Mathematics Subject Classification: 62J05, 62H12,15A06, 15A15.

1 Introduction

Linear algebraic methods have long played an essential role in the theory and teaching of statistics. The concepts in multivariate statistical analysis are nearly impossible to express without employing matrix notation and proper linear algebraic methods. These methods are especially important in branches of theoretical and applied statistics that exploit linear constructions to study empirical

**Corresponding author: E-mail: vpanov@ecko.uran.ru*

phenomena, and the theory of linear statistical models is therefore central to these areas of statistics [1]-[4].

Although the theory of linear statistical models was initially developed as a direct generalization of linear regression, its use was soon extended to a broader range of statistical methods, such as ANOVA and ANCOVA. Nevertheless, regression persists as the prototypical example of linear statistical models, and linear statistical models are sometimes referred to as general regression models [5]. The majority of the concepts and methods of linear statistical modelling appeared first in the context of regression and were later reformulated within a more general linear theory framework.

Let us first review some of the basic notions of linear statistical modelling [1]-[4]. The general equality specifying the linear model can be expressed as follows

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{Y} = [y_1, y_2, \ldots, y_n]'$ is an $n \times 1$-vector of observations of the dependent variable (the response Y), $\mathbf{X} = [x_{ij}]_{n\times(p+1)}$ is the matrix of observed values of the predictors (the design or model matrix), $\boldsymbol{\beta}$ is a $(p+1) \times 1$-vector of the model parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$-vector of unobserved residuals. The model matrix $\mathbf{X}$ can be presented as a row of columns $\mathbf{x}_j = [x_{ij}]_{n\times 1}$ of the observed values of the predictors $X_1, \ldots, X_p$, that is $\mathbf{X} = [\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p]$.

For instance, if the design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$

then the linear model represents a linear regression with p predictors, assuming that $x_{i0} = 1$:

$$y_i = \sum_{j=0}^{p} \beta_j x_{ij} + \varepsilon_i$$

Note that a multiple regression may include nonlinear dependence on predictors; e.g., if $x_{ij} = x_i^j$, then one obtains polynomial regression: $\mathbf{x}_j = \mathbf{x}^j,\ j = 0, 1, \ldots, p$.

The basic idea of linear modelling was adopted long ago in applied statistics; the first linear models were presented in geometric terms by Gauss and Legendre. The methodology of linear modelling was formalized in the middle of the twentieth century. The generalized linear model was proposed in 1972 by Nelder and Wedderburn [6], providing a unifying framework for many of the most commonly used statistical techniques (see, for example, [7] and [8]).

The least squares method provides a natural way to determine (or estimate) the values of the unknown parameters of a linear model. In this method, the sum over all observations of the squared errors

$$\arg \min_{\beta_0,\beta_1,\ldots,\beta_p} \sum_{j=0}^{n} \left(y_j - \sum_{i=0}^{p} \beta_i x_{ij} \right)^2 = \arg \min_{\beta_0,\beta_1,\ldots,\beta_p} \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \right)' \left(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \right), \tag{1.2}$$

is minimized. A common assumption employed in the linear model is that the errors are unbiased, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$. Without this assumption, estimates obtained for the parameters are only optimal in the least squares sense, and form the *best linear predictor* (BLP) or *best linear estimator* (BLE).

The BLP have many applications in theoretical statistics, but are less widely used than the *best linear unbiased predictor* (BLUP or BLUE) in applications. The BLUP is in fact equivalent to regression analysis for a normal (Gaussian) population [9]-[11]. For instance, in [10] and [12], [13] and some other books, regression is defined as the best linear predictor on the basis of a given set of predictors, rather than as a conditional response distribution with respect to this set of predictors. Hereafter, we will assume only that the model is linear and adopt no assumptions regarding the probability distributions of given variables and errors. In this case, the linear model expresses a linear relationship between the response Y and predictors $\{X_0 = 1,\ X_1, \ldots,\ X_p\}$, and is not related to the accuracy or power of the relationship.

2 Background and Related Studies

The problem of the relationship between models with different sets of predictors is not often discussed in the many books devoted to the general linear model and its applications (see, [1]-[3], [14], [15]). This problem is easily solvable if the predictor sets are connected by a linear nonsingular transformation (see lemma 5.2.1 below). The problem of determining new regression coefficients when independent variables are added or omitted was posed by Cochran in 1938 for a single regressor [16], and was later studied in the general case in [17]. The formulae obtained for the new coefficients used the observed values of the regressors and response, which was not convenient for establishing the relationships between the coefficients of the considered models. These formulae are presented in [18]-[20].

Explicit expressions for the dependencies between the parameters of regression models (for continuous variables) were obtained in [9] and [21] as a collateral application of the matrix SWEEP-operator. However, the proof of the formula is based on a highly specialized result from linear and computational algebra. In addition, the variables must be transformed so that they are mean-centered to apply SWEEP-operator, excluding the intercept term from the consideration in the regression model.

The explicit formulae under consideration have therefore had few applications, although they have been known for a long time. They may be of use in comparisons of regression models of different dimensions, but most authors have employed alternative methods to solve this problem (see, e.g., recent articles [22]-[27]. The theory of hierarchical (nested, multilevel) linear models is another area in which the above-mentioned formulae may be useful (see, [28] and [29]), but they have not yet been applied in it.

The need to study the relationships between regression models has arisen in biostatistics and epidemiology, primarily in the context of accounting for confounders [30], [31]. In general, the problem may be presented as follows. There is a set of independent variables that have an influence on the response and may be correlated with one another. A researcher is usually interested in just a few variables, often one or two, which influence the response that he or she wishes to study. The nature of individuals involved in an experiment may require certain variables other than the main variable(s) under study to be accounted for. Examples in medicine and epidemiology include the *age* and *gender* of the subjects (among others specific characteristics such as their ethnicity, personality traits, and abilities as well as the experimental conditions). Such mandatory independent variables are called *confounders* in epidemiology because they confound the influence of the main variable(s). A multiple model is therefore required, even though the researcher may only be interested in a few main variables. In implicit form, we have two models that must be compared, with and without confounders. For further discussion on confounders and their treatment, the reader is referred to [30]-[37].

The most recent direct application of the expressions related to Cochran's formulae appears in the context of the triangular systems theory [38] -[41]. Vellaisamy and Vijay have employed these

expressions to study the collapsibility of regression coefficients [42], see also [33]; and in [43], they were used to estimate the sensitivity of the confidence limits on the regression coefficients to the omission of a confounder.

In this paper, we do not consider specific statistical aspects of the dependency of linear statistical model parameters on different sets of predictors. Instead, we concentrate on the algebraic relationship between the parameters of various linear statistical models. An elementary proof of the theorem describing this dependence is obtained (Theorem 4.1). In contrast to other analogous statements, we work with original variables rather than mean-centered ones, so that the intercept terms are included in the consideration as well. We derive several corollaries of the theorem that have significance in their own right; one of them proposes a new interpretation of linear model parameters regardless of the correlation among the predictors.

3 Basic Notation

Let us consider a linear model with a single response Y and p predictors $X_1, \ldots, X_p$

$$\mathbf{Y} = \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\varepsilon}, \tag{3.1}$$

where the matrix $\mathbf{X}_p$ has dimensions $n \times (p+1)$. From the set of all predictors, we choose a subset $X_0, X_1, \ldots, X_q,\ q < p$ and consider the corresponding linear model

$$\mathbf{Y} = \mathbf{X}_q \boldsymbol{\alpha}_q + \boldsymbol{\varepsilon}, \tag{3.2}$$

where $\mathbf{X}_q$ is of order $n \times (q+1)$. The model (3.2) may be considered to be obtained from model (3.1) by excluding $p-q$ predictors. Conversely, the model (3.1) is obtained from (3.2) by adding predictors $X_{q+1}, \ldots, X_p$. The main goal of this paper is to prove Theorem 4.1 (below) and illustrate some of its applications. Theorem 4.1 states that there is a linear relationship between models (3.1) and (3.2) in terms of their parameters (i.e. their estimates) rather than the design matrix. The suggested proof is elementary and accessible to non-specialists in linear algebra. Let us introduce the auxiliary linear models

$$\mathbf{x}_i = \mathbf{x}_q \boldsymbol{\gamma}_{iq} + \boldsymbol{\varepsilon}, \tag{3.3}$$

where

$$\boldsymbol{\gamma}_{iq} = (\gamma_{i0}, \gamma_{i1}, \ldots, \gamma_{iq})'.$$

The matrices $\mathbf{B}_p$ and $\mathbf{A}_q$ consist of estimates of the model parameters of (3.1) and (3.2), respectively. The estimates for the parameters of model (3.3) are contained in the matrix $\mathbf{C}_{np}$. Explicitly, these matrices are given by

$$\mathbf{A}_q = [a_0, a_1, a_2, \ldots, a_q]_{1\times(q+1)} \tag{3.4}$$

$$\mathbf{B}_p = [b_0, b_1, b_2, \ldots, b_p]_{1\times(p+1)} \tag{3.5}$$

$$\mathbf{C}_{np} = [c_{ij}]_{(p+1)\times(q+1)}, \quad 0 \le i \le p,\ 0 \le j \le q, \tag{3.6}$$

assuming that $c_{ij} = \delta_{ij}$ for $i, j \in \{0, 1, \ldots, q\}$.

The estimates are obtained using the least squares method, by minimizing the sum of the squared errors over the observations for each linear model. For model (3.1), this problem has been stated above in (1.2). The minimization problem for the other models can be similarly formulated.

Solutions to the optimization problem (1.2), which are estimates of the parameters $\boldsymbol{\beta}_p,\ \boldsymbol{\alpha}_q,\ \boldsymbol{\gamma}_{iq}$, may be found from the following systems of linear equations

$$\mathbf{B}_p \cdot \mathbf{X}\mathbf{X}_p = \mathbf{Y}\mathbf{X}_p \tag{3.7}$$

$$\mathbf{A}_q \cdot \mathbf{XX}_q = \mathbf{YX}_q, \tag{3.8}$$

$$(c_{i0}, c_{i1}, \ldots, c_{iq}) \cdot \mathbf{XX}_q = \mathbf{XX}_{iq}, \tag{3.9}$$

where the index q denotes the use of the chosen set of predictors $\{X_0, X_1, \ldots, X_q\}$, and the following matrices have been introduced (where a bar denotes the mean value):

$$\begin{gathered}
\mathbf{XX}_p = \left(\overline{X_i X_j}\right)_{(p+1)\times(p+1)} \quad i,j = 0,1,\ldots,p \\
\mathbf{XX}_q = \left(\overline{X_i X_j}\right)_{(q+1)\times(q+1)} \quad i,j = 0,1,\ldots,q \\
\mathbf{YX}_p = \left(\overline{Y X_i}\right)_{1\times(p+1)}, \quad i = 0,1,\ldots,p \\
\mathbf{YX}_q = \left(\overline{Y X_i}\right)_{1\times(q+1)}, \quad i = 0,1,\ldots,q \\
\mathbf{XX}_{iq} = \left(\overline{X_i X_j}\right)_{1\times(q+1)}, \quad i = 0,1,\ldots,p,\ j = 0,1,\ldots,q
\end{gathered}$$

4 Main Theorem

Let us now state and prove the main theorem.

Theorem 4.1. *The following equality holds:*

$$a_j = \sum_{i=0}^{p} b_i c_{ij}, \quad j = 0,1,\ldots,q, \tag{4.1}$$

or

$$\mathbf{A}_q = \mathbf{B}_p \cdot \mathbf{C}_{pq}. \tag{4.2}$$

Proof. Let $\mathcal{H}$ denote a real Hilbert space and $\mathcal{H}_n$ its n-dimensional subspace, and $\mathcal{H}_k$-k-dimensional subspace in $\mathcal{H}_n$, $k \le n$. There are the following projection operators: P_n is a projection from $\mathcal{H}$ onto the space $\mathcal{H}_n$, P_k is a projection from $\mathcal{H}$ onto the space $\mathcal{H}_k$, and $P_{n,k}$ is a projection from $\mathcal{H}_n$ onto its subspace $\mathcal{H}_k$. Then for any $\mathbf{x} \in \mathcal{H}$ the difference $\mathbf{x} - P_n\mathbf{x} \perp \mathcal{H}_n$. Thus, $\mathbf{x} - P_n\mathbf{x} \perp \mathcal{H}_k$ and

$$P_k\mathbf{x} = P_k\left((x - P_n\mathbf{x}) + P_n\mathbf{x}\right) = P_k\left(\mathbf{x} - P_n\mathbf{x}\right) + P_kP_n\mathbf{x} = P_kP_n\mathbf{x} = P_{n,k}P_n\mathbf{x}$$

which proves the theorem. Indeed, let us take $\mathcal{H} = \mathcal{H}_n$ to be the span of vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_p$, i.e. $\mathcal{H}_n = \langle \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_q \rangle$, and $\mathcal{H}_k = \langle \mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_q \rangle$, so we have $n = p+1$, $k = q+1$. Then for any $\mathbf{x} \in \mathcal{H}$ the following equalities hold

$$\mathbf{x} = \sum_{i=0}^{p} b_i\mathbf{x}_i, \quad P_k\mathbf{x} = \sum_{j=0}^{q} a_j\mathbf{x}_j, \quad P_{n,k}\mathbf{x}_i = \sum_{j=0}^{q} c_{ij}\mathbf{x}_j,$$

which result in equality (4.2). In these notations, the equality (4.2) can be presented as the commutative diagram

$$\begin{array}{ccc}
\mathcal{H} & = & \mathcal{H} \\
{\scriptstyle P_n}\downarrow & & \downarrow{\scriptstyle P_k} \\
\mathcal{H}_n & \xrightarrow[P_{n,k}]{} & \mathcal{H}_k
\end{array}$$

that is

$$P_k = P_{n,k} \circ P_n$$

□

Remark 4.1. It is evident that Theorem 4.1 holds regardless of the order of the predictors in the selected set. One just needs more complex notation for the indices. Moreover, Theorem 4.1 remains true for a multivariate response $(Y_1, Y_2, \ldots, Y_k)$, if we employ component-wise least squares. In the next remark, we use such notation to introduce the general form of Theorem 4.1.

Remark 4.2. In the following, we sometimes use more compact notations for equality (4.2). We owe these notations to Yule and they are widely used in research on regression analysis (see, for instance, [41], [44], [45]). In general, a set of random variables $(Y_1, Y_2, \ldots, Y_m)$ can be partitioned into disjointed sets denoted by a, b and c. The coefficients of the BLE of system a on system b are denoted by $\Pi_{a|b}$ and the notations $\Pi_{a|bc}$ and $\Pi_{bc|b}$ are defined similarly, where bc is a union of systems b and c. For all cases, the BLP includes the constant predictor $\mathbf{X}_0 = [1, 1, \ldots, 1]'$. It is straightforward to show that Theorem 4.1 holds for the multivariate response $\mathbf{Y}_a = \{Y_i\}_{i \in a}$ as well. The equality (4.2) can therefore be expressed as follows

$$\Pi_{a|b} = \Pi_{a|bc}\Pi_{bc|b}.$$

The matrix $\Pi_{a|bc}$ can be expressed as the concatenation of the two submatrices generated by systems b and c:

$$\Pi_{a|bc} = \begin{bmatrix} \Pi_{a|b.c} & \Pi_{a|c.b} \end{bmatrix},$$

where $\Pi_{a|b.c}$ is the matrix of the coefficients from $\Pi_{a|bc}$ that are multiplied by the variables from b in the linear model with the complete set of predictors, bc. The matrix $\Pi_{a|c.b}$ has similar definition. It is readily understood that matrix $\Pi_{bc|b}$ may be expressed in the simpler form

$$\Pi_{bc|b} = \begin{bmatrix} \mathbf{I} \\ \Pi_{c|b} \end{bmatrix},$$

where $\mathbf{I}$ is the identity matrix whose order is one higher than the number of predictors in system b. Equality (4.2) can then be expressed as follows:

$$\Pi_{a|b} = \Pi_{a|b.c} + \Pi_{a|c.b}\Pi_{c|b}.$$

This expression is widely used in some approaches to the study of statistical dependencies, such as the theory of triangular systems and graph chains [39], [41], [46]-[48].

In the case under consideration, the total set of predictors is formed by the response and all of the independent predictors, i.e.,

$$abc = \{Y, X_1, X_2, \ldots, X_p\},\ a = \{Y\},\ b = \{X_1, \ldots, X_q\},\ c = \{X_{q+1}, \ldots, X_p\}.$$

The matrices introduced in the preceding paragraph are then as follows

$$\begin{aligned} \Pi_{a|b} &= [a_0, a_1, \ldots, a_q] & \dim\left(\Pi_{a|b}\right) &= 1 \times (q+1) \\ \Pi_{a|bc} &= [b_0, b_1, \ldots, b_p] & \dim\left(\Pi_{a|bc}\right) &= 1 \times (p+1) \\ \Pi_{bc|b} &= [c_{ij}] & \dim\left(\Pi_{bc|b}\right) &= (p+1) \times (q+1). \end{aligned}$$

5 Some Applications of the Main Theorem

We now discuss some consequences of Theorem 4.1. First, we briefly examine the simple regression.

5.1 Relationship between the Multiple and Simple Regressions

Let us consider a set of p predictors, $X_1, X_2, \ldots, X_p$, and a single response variable, Y. The multiple regression equation for the response Y on the set of all of the given predictors (including $X_0 = 1$) has the form

$$y = \sum_{i=0}^{p} b_i x_i.$$

Let us introduce a simple regression for any given predictor on another predictor

$$x_i = c_{ij0} + c_{ij}x_j, \quad i, j = 0, 1, \ldots, p,$$

assuming that $c_{ii0} = 0,\ c_{ii} = 1$.

We also need simple regressions of the response on each predictor. The equations of these regressions are

$$y = a_{i0} + a_i x_i, \quad i = 0, 1, \ldots, p$$

In the case under consideration, we have $q = 1$, and each predictor is taken to be a one-element predictor set. According to Theorem 4.1, for every $i \in \{0, 1, \ldots, p\}$ we therefore have the equality

$$a_j = \sum_{i=0}^{p} b_i c_{ij}. \tag{5.1}$$

Finally, we introduce the following matrices:

$$\mathbf{A} = [a_0, a_1, \ldots, a_p]_{1\times(p+1)}$$
$$\mathbf{B} = [b_0, b_1, \ldots, b_p]_{1\times(p+1)}$$
$$\mathbf{C} = [c_{ij}]_{(p+1)\times(p+1)}.$$

Equality (5.1) can then be presented as the single matrix equality

$$\mathbf{A} = \mathbf{B} \cdot \mathbf{C}, \tag{5.2}$$

as is derived in [49].

As a_0 is a predicted value of Y at $\mathbf{X} = [\mathbf{1} \quad \mathbf{0}]$, we obtain an expression for a_0 in terms of regression parameters:

$$a_0 = \sum_{i=0}^{p} b_i c_{i0}.$$

5.2 Multiple Cochran's Formula

Cochran's problem of change in the predictor set was mentioned in the Introduction. Originally, this problem was posed for regression models, but the formulation for the general linear model was straightforward. Initially, a single variable was included or excluded from a given multiple regression model. From Theorem 4.1, one can readily obtain a generalization of Cochran's formula for a set of included or excluded predictors in a linear model. First, it is clear that Theorem 4.1 is itself a case of omitting variables.

Theorem 5.1. *(Exclusion of predictors). If a subset $X_1, X_2, \ldots, X_q$ is chosen from the set $X_1, X_2, \ldots, X_p$ of explanatory variables, then the coefficients of the multiple regression equation for the response Y with the variable set $X_1, X_2, \ldots, X_q$ are expressed through the coefficients of the multiple regression equation for the response Y with the variable set $X_1, X_2, \ldots, X_p$ by equality* (4.2).

Applications and discussion of omitting variables in different areas are considered in many books and articles, some recent ones are [50]-[52].

An analogous statement for the addition of predictors may have different formulations, depending on *how* the variables are added. For instance, one can add all of the supplementary variables simultaneously or add them one by one, recalculating the model coefficients each time a variable is added. For the one-stage addition, the corresponding statement is

Theorem 5.2. *(Addition of predictors). Let $X_0 = 1, X_1, \ldots, X_p$ be a set of explanatory variables, let $X_0, X_1, \ldots, X_q$ be the chosen set of predictors and let Y be the response. Let us introduce the following regression equations ($x_0 = 1$)*

$$y = \sum_{i=0}^{p} b_i x_i, \qquad y = \sum_{j=0}^{q} a_j x_j, \quad y = a'_0 + \sum_{j=q+1}^{p} a_j x_j$$

$$x_i = \sum_{j=0}^{q} c_{ij} x_j, \quad i = 0, 1, \ldots, p, \quad c_{ii} = 1$$

$$x_i = d_{i0} + \sum_{j=q+1}^{p} d_{ij} x_j \quad i = 0, 1, \ldots, p, \quad d_{ii} = 1.$$

Define the matrices $\mathbf{A},\ \mathbf{B}_p$ and $\mathbf{C}$

$$\mathbf{A} = [a_0, a_1, \ldots, a_p], \quad \mathbf{B}_p = [b_0, b_1, \ldots, b_p]$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_{q+1} & D_{p-q} \\ C_q & \mathbf{I}_{p-q} \end{bmatrix},$$

where

$$C_q = [c_{ij}], \quad i = q+1, \ldots, p, \quad j = 0, 1, \ldots, q,$$
$$D_{p-q} = [d_{ij}] \quad i = 0, 1, \ldots, q,\ j = q+1, \ldots, p,$$

and $\mathbf{I}_m$ is the identity matrix of order m. From Theorem 4.1, we then obtain the equality

$$\mathbf{A} = \mathbf{B}_p \cdot \mathbf{C},$$

and if the matrix $\mathbf{C}$ is nonsingular, we obtain

$$\mathbf{B}_p = \mathbf{A} \cdot \mathbf{C}^{-1}.$$

This equality shows how the coefficients of a linear regression model change in response to the addition of a set of predictors. To obtain new coefficients, it would apparently suffice to know the regression coefficients of the initial model (row $[a_0, a_1, \ldots, a_k]$), the regression coefficients of the model based on the added set of predictors (row $[a_{k+1}, \ldots, a_p]$), and the regression matrices of a new system of predictors based on the old one and those of the old system of predictors based on the new one (matrices C_k and D_{n-k}). If the matrix $\mathbf{C}$ is singular, then one could use a generalized inverse or choose another set of additional variables to construct the matrix $\mathbf{C}$.

Remark 5.1. As $\mathbf{C}$ is a block matrix, its inverse can be obtained via the Schur complement $\mathbf{C} \mid E_{q+1}$ [53], [54]. We have

$$\mathbf{C} \mid E_{q+1} = E_{p-q} - \mathbf{C}_q \cdot D_{p-q}.$$

In the simplest case of $p = 2,\ q = 1$, we obtain

$$C_1 = c_{21}, \quad D_1 = c_{12},$$

where c_{21} and c_{12} are coefficients in the linear regression models

$$x_2 = c_{210} + c_{21} x_1 \quad x_1 = c_{120} + c_{12} x_2.$$

We therefore have the equality

$$[b_1, b_2] = [a_1, a_2] \cdot \begin{bmatrix} 1 & c_{12} \\ c_{21} & 1 \end{bmatrix}^{-1}.$$

In the case under consideration, there is the well-known equality $c_{12} \cdot c_{21} = r^2$, where r is Pearson's correlation coefficient between the predictors X_1, X_2. From Theorem 4.1 and the previous formula in which the inverse matrix is calculated, we obtain the following formulae, which show how simple regression coefficients can be calculated from multiple regression coefficients and vice versa

$$a_1 = b_1 + b_2 c_{21}, \quad a_2 = b_2 + b_1 c_{12}$$
$$b_1 = \frac{a_1 - a_2 c_{21}}{1 - r^2}; \quad b_2 = \frac{a_2 - a_1 c_{12}}{1 - r^2}. \tag{5.3}$$

From these expressions, one readily obtains corresponding equalities for the standardized regression coefficients $b'_k = b_k \cdot \frac{s_{x_k}}{s_y}, k = 1, 2$, where s_z is the standard deviation of the variable $Z \in \{X_k, Y\}$

$$b'_1 = \frac{r_{y1} - r_{y2} r}{1 - r^2}, \quad b'_2 = \frac{r_{y2} - r_{y1} r}{1 - r^2}.$$

Here we use the fact that the standardized coefficient a'_k for univariate regression of the response Y on the predictor X_k is simply the Pearson's regression coefficient between the outcome and predictor, $a'_k = r_{yk}$.

Finally, the following expressions represent the regression coefficients between the predictors in terms of those for the simple and multiple regression models for outcome Y, and follow easily from the previous formulae

$$c_{12} = \frac{a_2 - b_2}{b_1} \quad c_{21} = \frac{a_1 - b_1}{b_2}.$$

All of these formulae can be generalized to an arbitrary number of predictors. These generalizations are known, but their proofs are typically more difficult than those presented above (see, e.g., [55]).

These equalities can be used to analyze the impact of additional variables on the main predictor in simple linear model. Consider model with a single predictor X_1,

$$y = a_{10} + a_1 x_1, \tag{5.4}$$

and suppose that we are investigating a linear model with two predictors X_1 and X_2. The new variable, X_2, accounts for the new circumstances of the experiment and may have influence on the outcome Y and the predictor X_1. Therefore, it may change the coefficient a_1 in the univariate model (5.4) into the coefficient b_1 in the bivariate model

$$y = b_0 + b_1 x_1 + b_2 x_2. \tag{5.5}$$

From Theorem 4.1, the difference between the parameters a_1 and b_1 is

$$\Delta a_1 = a_1 - b_1 = c_{21} \cdot b_2.$$

This represents two causes of the influence of the variable X_2 being added to the univariate model (5.4). The first cause is the correlation between the variables X_1 and X_2, expressed by the parameter c_{21}, and the second cause is the impact of X_2 on the outcome Y, expressed by b_2. It is important to understand that the parameters c_{21} and b_2 are not independent if the variables X_1 and X_2 are not independent. Therefore, the influence of X_2 cannot be strictly separated into independent causes. One can only state conditions under which the increment Δa_1 is non-zero, given by the simultaneous inequalities

$$c_{21} \neq 0, \ b_2 \neq 0$$

Similar calculations can be performed in the general case of p predictors, although the corresponding interpretation is less straightforward. In this case, the change in the coefficient a_1 after adding the variables $X_2, \ldots, X_p$ to the univariate model (5.4) is

$$\Delta a_1 = a_1 - b_1 = \sum_{i=2}^{p} b_i \cdot c_{i1}.$$

5.3 Interpretation of Linear Model Parameters

The interpretation of linear regression model parameters is an old and relatively straightforward problem. In theoretical statistics, it is well known that no interpretation is possible for the parameters of a multivariate regression model when the regressors are correlated [25], [56], [57]. The only case in which such an interpretation is valid is where the predictors are independent of one another. In that case, each predictor can change with no impact on the other predictors, and the coefficient b_i in the multivariate model

$$y = \sum_{i=0}^{p} b_i x_i$$

is equal to the change in the mean value of the outcome under a 1-unit increment in the variable X_i (holding all of the other variables constant). The magnitude of the coefficient therefore indicates the contribution of the variable X_i to the variation of the outcome. It is important to emphasize that this interpretation is valid only for uncorrelated predictors. If there is a significant correlation between the independent variables, then no consistent interpretation of the model parameters has yet been proposed.

Nevertheless, there is an alternative way to interpret the linear model parameters, based on geometric reasoning and Theorem 4.1. Let us return to the multivariate model (3.1) and introduce a variable $X_1^\star$, defined by the equality

$$X_1^\star = X_1 - \sum_{i=0, i\neq 1}^{p} c_{1i} X_i, \tag{5.6}$$

where $[c_{1i}]$ is the first row of the matrix $\mathbf{C}_{np}$ (see (3.4)). A similar variable appears in implicit form in some other regression/correlation analysis formulations [44], [58]. We will return to the interpretation of this variable later on. In addition to the multivariate linear model (3.1), we consider the univariate model

$$\mathbf{Y} = \alpha_0^\star + \alpha_1^\star \mathbf{x}_1^\star + \boldsymbol{\varepsilon},$$

where $\mathbf{x}_1^\star$ is the column of observed values of the variable $X_1^\star$.

The following theorem is well-known [10], [12] and its proof can be obtained as a consequence of Theorem 4.1.

Theorem 5.3. *If a random variable X_1 is uncorrelated with the response Y, and with each remaining predictor $X_2, \ldots, X_p$, then the estimate b_1 of the parameter β_1 in the model*

$$\mathbf{Y} = \sum_{i=0}^{p} \beta_i \mathbf{x}_i + \boldsymbol{\varepsilon}.$$

equals zero: $b_1 = 0$.

The main statement in this section is Theorem 5.4, which may be viewed as a generalization of Theorem 5.3.

Theorem 5.4. *Let*

$$\mathbf{Y} = \mathbf{X}_p \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

be the linear model for the response Y on p predictors $X_1, \ldots, X_p$, with $X_0 = 1$, and the model (index in round brackets is omitted)

$$\mathbf{x}_1 = c_{10.2\ldots p} + \sum_{i=2}^{p} c_{1i.2\ldots(i)\ldots p} \mathbf{x}_i + \boldsymbol{\varepsilon} \tag{5.7}$$

expresses the linear dependence of the predictor X_1 on the remaining predictors. Let us introduce the variable

$$X_1^\star = X_1 - \sum_{i=2}^{p} c_{i.2\dots(i)\dots p} X_i - c_{10.2\dots p}, \tag{5.8}$$

and consider the univariate linear model for the response Y on the predictor $X_1^\star$,

$$\mathbf{Y} = a_0^\star + a_1^\star \mathbf{x}_1^\star.$$

Then,

$$b_1 = a_1^\star.$$

Let us denote the sequence $23\dots p$ with omitted number i by $q_i = 2\dots(i)\dots p$ and let q_0 denote the complete sequence $23\dots p$. Then, model (5.7) takes the form

$$\mathbf{x}_1 = c_{10.q_0} + \sum_{i=2}^{p} c_{1i.q_i} \mathbf{x}_i + \boldsymbol{\varepsilon}.$$

Proof will be divided into a sequence of lemmas.

Lemma 5.5. *Let the variable X_1 be transformed as follows*

$$X_1^\star = X_1 - \sum_{i=2}^{p} \gamma_i X_i - \gamma_0.$$

Then, estimates of the parameters $\boldsymbol{\beta}^\star$ of the linear model with predictors $X_0, X_1^\star, \dots, X_p$ are obtained by the matrix multiplication

$$\boldsymbol{\beta}^\star = \Gamma^{-1} \cdot \boldsymbol{\beta},$$

where Γ is the transformation matrix, i.e.,

$$\Gamma = \begin{bmatrix} 1 & -\gamma_0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -\gamma_2 & 1 & \dots & 0 \\ 0 & -\gamma_3 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & -\gamma_p & 0 & \dots & 1 \end{bmatrix}_{(p+1)\times(p+1)},$$

and the elements of $\boldsymbol{\beta}$ are estimates of the parameters of the linear model with predictors $X_0, X_1, \dots, X_p$.

Proof. The lemma is a particular case of the following statement. The proof is given only for completeness.

Let a linear transformation of the response and predictors be given by the formulae

$$\widetilde{\mathbf{Y}} = \alpha \mathbf{Y} + \mathbf{X}\mathbf{a}, \quad \widetilde{\mathbf{X}} = \mathbf{X}\mathbf{A},$$

where α is a scalar, $\mathbf{a}$ is a vector and $\mathbf{A}$ is a non-singular square matrix. Letting $\boldsymbol{\beta}$ denote the estimates of the parameters of the linear model with the predictor set $X_0, X_1, \dots, X_p$ and $\widetilde{\boldsymbol{\beta}}$ denote the estimates on the predictor set $\widetilde{\mathbf{X}}$, we obtain the following from the normal equations for the estimates [10], [12], [56]:

$$\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}\right)^{-1} \widetilde{\mathbf{X}}^T \widetilde{\mathbf{Y}} = \left(\mathbf{A}^T \mathbf{X}' \mathbf{X} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{X}^T (\alpha \mathbf{Y} + \mathbf{X}\mathbf{a}) =$$

$$= \mathbf{A}^{-1} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T (\alpha \mathbf{Y} + \mathbf{X}\mathbf{a}) =$$

$$= \mathbf{A}^{-1} \left(\alpha \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{Y} + \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{X}\mathbf{a} \right) =$$

$$= \mathbf{A}^{-1} (\alpha \boldsymbol{\beta} + \mathbf{a}).$$

In our case, $\alpha = 1, \mathbf{a} = 0, \mathbf{A} = \Gamma$, and we therefore obtain the formula

$$\widetilde{\boldsymbol{\beta}} = \Gamma^{-1}\boldsymbol{\beta}.$$

By direct calculation, we can check that

$$\Gamma^{-1} = \begin{bmatrix} 1 & \gamma_0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & \gamma_2 & 1 & \dots & 0 \\ 0 & \gamma_3 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \gamma_p & 0 & \dots & 1 \end{bmatrix}.$$

The following equalities therefore hold:

$$\widetilde{\boldsymbol{\beta}} = \begin{bmatrix} \widetilde{\beta}_0 \\ \widetilde{\beta}_1 \\ \widetilde{\beta}_2 \\ \vdots \\ \widetilde{\beta}_p \end{bmatrix} = \begin{bmatrix} \beta_0\gamma_0 \\ \beta_1 \\ \beta_3 + \gamma_2\beta_1 \\ \vdots \\ \beta_p + \gamma_p\beta_1 \end{bmatrix}.$$

In particular, $\widetilde{\beta}_1 = \beta_1$. Therefore, making a change in the predictors by replacing X_1 with $X_1^\star$ does not change the coefficient for the predictor X_1. □

The next lemma is well-known [10], [20], [56].

Lemma 5.6. *Let Y be a response and let $\{X_0 = 1, X_1, \dots, X_p\}$ be a predictor set. Let $\boldsymbol{\varepsilon}_Y = Y - \sum_{i=0}^{p} b_{Yi.q_i} X_i$ be a residual of the response Y for the linear model with predictors $\{X_0 = 1, X_1, \dots, X_p\}$. Then, $\boldsymbol{\varepsilon}_Y$ has zero correlation with each predictor.*

Lemma 5.6 is used below as follows. Consider the predictor X_1 as a response and a set of explanatory variables consisting of all of the remaining predictors. Then, the variable $X_1^\star = X_1 - \sum_{i=0, i\neq 1}^{p} c_{1i,q_i} X_i$ has zero correlation with every predictor in the set $\{X_2, \dots, X_p\}$. In addition, it is straightforward to show that $X_1^\star$ is uncorrelated with X_1 [56].

Take the systems a, b and c described in Theorem 5.3, with X_1 replaced by $X_1^\star$. Estimates of the parameters for the model with the new set of predictors will be denoted by a superscript asterisk. From Lemma 5.6 it follows that $\{X_2, \dots, X_p\}$ are not correlated with $X_1^\star$. Therefore,

$$\begin{aligned} \Pi_{a|b} &= [a_0^\star, a_1^\star] \\ \Pi_{a|b.c} &= [b_0^\star, b_1^\star] \\ \Pi_{a|c.b} &= [b_2^\star, \dots, b_p^\star] \\ \Pi_{c|b} &= \begin{bmatrix} c_{20.q_2} & 0 \\ c_{30.q_3} & 0 \\ \vdots & \vdots \\ c_{p0.q_p} & 0 \end{bmatrix}. \end{aligned}$$

In the case under consideration, Theorem 4.1 implies that

$$\Pi_{a|b} = \Pi_{a|b.c} + \Pi_{a|c.b}\Pi_{c|b}.$$

We therefore obtain $b_1^\star = a_1^\star$, and it follows by Lemma 5.5 that $b_1 = b_1^\star$. Therefore,

$$b_1 = a_1^\star.$$

Theorem 5.4 provides a new interpretation of linear model parameters, as follows. As has been mentioned above, in the univariate linear model

$$y = a_0 + a_1 x,$$

the slope a_1 represents a change in the mean value of the outcome Y under a 1-unit increment in the variable X. From Theorem 5.4, it follows that the coefficient b_1 of the variable X_1 in the linear model (3.1) represents a change in the mean value of the outcome under a 1-unit increment in the combined variable $X_1^\star$ defined in Lemma 5.5. The interpretation of the linear model parameter b_1 therefore shifts to the interpretation of a new variable, $X_1^\star$. A similar linear combination of independent variables is well known in factor analysis, but in that case the orthogonalization process generates usually uninterpretable variables. The $X_1^\star$ can be interpreted as the portion of the variable X_1 from which the linear contribution of the other independent variables is excluded. Alternatively, in regression terms, the variable $X_1^\star$ is the portion of X_1 that cannot be explained via linear regression on the other regressors.

6 Applications to Real Data Analysis

6.1 Regression with two Predictors

Let us consider the use of Theorem 5.4 for investigating the dependency of incidence on various air pollution toxicants of City St.-Petersburg (Russia). The primary data were published in [59]. In the remainder of this section, we assume incidence to be incidence rate in the adult population (i.e. the number of disease cases per 1000 adult population a year) averaged over a 5-year observation period. In the primary data, the rates of incidence were gathered across 19 boroughs of St.-Petersburg. We consider toxicant concentrations as random variables, i.e. mean toxicant concentration expressed in maximum concentration limit (MCL) terms and averaged over 5-year observation period. Each of these variables takes on 19 values in accordance with the number of boroughs. We denote these covariates by the usual chemical notations: CO, NO_2, SO_2, Pb etc. (the data consists of 12 pollutants).

The simple linear regression equations of response Y(incidence) on concentrations of CO and NO_2 are given by

$$Y = 603 + 579\,CO \tag{6.1}$$

$$Y = 414 + 416\,NO_2 \tag{6.2}$$

According to equation (6.1), incidence increases by 579 cases per 1000 population at an increase in CO concentration by MCL unit a year. Equation (6.2) may be interpreted in the same way. In short, both CO and NO_2 increase incidence.

There is a tight positive correlation between predictors CO and NO_2. Pearson's correlation coefficient is 0.75, and the regression equation is

$$CO = -0.131 + 0.576\,NO_2$$

This shows that growth in one toxicant is related to growth in another. Hence, one can conjecture that equation (6.1) does describe an increase in incidence at a simultaneous increase in both pollutants (CO and NO_2). A question then arises: could one specify the 'pure' influence of each toxicant on incidence, separating the contribution of one toxicant from that of the other?

To extract the contribution of each toxicant to the incidence in the presence of other toxicants, researchers often use a multiple regression equation including all toxicants. Such interpretation is common in some biological and medical applications of regression analysis. We refer to [60] as a typical exposition. In the case under consideration, we obtain a multiple regression equation

$$Y = 465 + 390\,CO + 191\,NO_2 \tag{6.3}$$

A lot of authors consider the coefficients of a multiple regression equation obtained by means of the least squares method to be meaningless if there are correlations among predictors (see, e.g., [56], [57]). These coefficients cannot be used to assess separately the dependence of Y on CO and Y on NO_2. Nevertheless, there are other authors who treat each coefficient of a multiple regression equation as the contribution of an individual toxicant to the outcome against the background of other toxicants (e.g., [60]). Moreover, this contribution has to be refined as compared to (6.1)–(6.2). Their supposition is that predictors as if distribute their influence on the outcome in a multiple regression equation so that each predictor describes its influence with the other being in the background. According to this viewpoint, the addition of another toxicant, NO_2, to CO and change from (6.1) to (6.3) should attenuate the effect of CO because the corresponding coefficient diminished from 579 to 390. The same conclusion holds for NO_2 and CO and equations (6.2) and (6.3).

These authors do not provide any substantive explanation for the biomedical meaning of variations in the coefficients in (6.1)–(6.2) and (6.3); nor do they explain the refined contribution of each individual toxicant. Variations in regression coefficients could be explained by going over from simple regressions (6.1) or (6.2) to multiple regression (6.3). Indeed, coefficient $b_1 = 390$ in equation (6.3) is equal to coefficient a_1^* in the simple regression equation

$$Y = a_{10}^* + a_1^* CO^*,$$

where covariate CO^* is defined by

$$CO^* = CO - 0.576\,NO_2 \tag{6.4}$$

By (6.4), predictor CO^* is obtained from CO by excluding its part correlated with NO_2. Then $b_1 = 390$ means an increased incidence rate at a growth in CO concentration excluding the linear statistical dependence of CO and NO_2.

One can similarly treat coefficient $b_2 = 191$ in (6.3). It is equal to a_2^* in the simple regression equation

$$Y = a_{20}^* + a_2^*\,NO_2^*,$$

where NO_2^* is a part of toxicant NO_2 which contains no linear statistical dependence on CO.

We seem to have obtained a consistent picture: by excluding the (linear) dependence of one toxicant on the other we arrive at a 'pure' influence of a particular factor on incidence. Since both factors increase the incidence, and the concentration of each factor increases with growth in the other, one can anticipate that the magnitudes of the coefficients in equation (6.3) should be less than in (6.1)–(6.2). This is exactly so in the case under consideration.

It is not as simple as that though. Let us consider the dependence of incidence Y on the concentrations of CO and SO_2. A simple regression equation of Y on SO_2 is given by

$$Y = 919 + 52\,SO_2$$

The association between CO and SO_2 is very similar to that between CO and NO_2. For instance, the correlation coefficient is 0.73 and the regression equation is

$$CO = 0.272 + 0.316\,SO_2 \tag{6.5}$$

The multiple regression equation in the case considered is

$$Y = 634 + 1047\,CO - 278\,SO_2 \tag{6.6}$$

Assuming the coefficients of (6.6) to be refined ones we should treat the magnitude 1047 as a 'pure' influence of CO against the background of SO_2, and -278 as a 'pure' influence of SO_2 against the background of CO. Obviously, such interpretation of regression coefficients is invalid, since the 'pure' influence of toxicant SO_2 becomes negative. The reason for such misinterpretation is the tight correlation between predictors CO and SO_2. One has to take into account this correlation in treating regression coefficients.

The coefficient at CO in ((6.6) is twice as large as that in (6.1). By Theorem 5.4, coefficient $b_1 = 1047$ is equal to the slope in

$$Y = 697 + 1047\,CO^*, \tag{6.7}$$

where $CO^* = CO - 0.316\,SO_2$. In biomedical terms, we obtain an inexplicable picture: we have reduced the toxic burden on the population by removing one of the two toxicants, but the incidence grows with CO even more rapidly. In mathematical terms, we can explain this as follows. It is clear from the definition of CO^* that its range is less than the range of CO. In both cases, the incidence is the same, which implies an increase in coefficient b_1.

Generally, inequality $b_1 > a_1$ is impossible if we consider the multiple regression coefficients as refined ones. But if we refer to equality (5.3), we can see that under $a_2 \ll a_1$ and correlation coefficient r close to 1, inequality $b_1 > a_1$ may hold true. The formula (5.3) also explains the possibility of a negative value for coefficient b_2.

6.2 Regression with three Predictors

Let us consider a regression equation of incidence Y on three predictors CO, NO_2 and SO_2. By the least square method, we obtain the equation

$$Y = 494 + 857\,CO + 194\,NO_2 - 279\,SO_2$$

Equation (5.7) becomes

$$CO = -0.108 + 0.386\,NO_2 + 0.195\,SO_2$$

The new variable CO^* is defined by (5.8), and the simple regression equation for Y on this predictor is given by

$$Y = 1076 + 857\,CO^*$$

We see that $b_1 = a_1^*$ as well.

Note that the correlation coefficient of model (6.5) is $r = 0.74$, and that of model (6.7) is $r = 0.46$. The latter is less than the coefficient of correlation between incidence Y and CO ($r = 0.58$).

7 Conclusion

We have provided an elementary proof of several linear relationships between the parameters of linear statistical models. The statement of Theorem 4.1 is proven by an elementary reasoning. We represent the statement in a form more convenient for applications, using more explicit notations. We have devoted particular attention to deriving some consequences of Theorem 4.1. There may yet be additional interesting corollaries of Theorem 4.1 and other theorems besides those presented in this paper, which may further expand potential applications of the discussed relationships.

Competing Interests

The authors declare that no competing interests exist.

References

[1] Christensen R. Plane answers to complex questions. 4th ed. Springer New York; 2011.

[2] Clarke BR. Linear models. The theory and application of analysis of variance. John Wiley & Sons, Inc., Hoboken, NJ, USA; 2008.

[3] Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. 4rd ed. McGraw-Hill/Irwin, New York, NY; 2005.

[4] Rao CR, Toutenburg H. Linear models: Least squares and alternatives. 2nd ed. Springer; 1999.

[5] Kariya T, Kurata H. Generalized least squares. John Wiley & Sons; 2004.

[6] Nelder JA, Wedderburn RWM. Generalized linear models. J Royal Stat Soc, Series A. 1972; 135(3):370-384.

[7] Dobson AJ. An introduction to generalized linear models. 2nd ed. Chapman & Hall/CRC, Boca Raton; 2002.

[8] McCullagh P, Nelder JA. Generalized linear models. 2nd ed. Chapman & Hall/CRC, London; 1989.

[9] Dempster AP. Elements of continuous multivariate analysis. Addison-Wesley Pub. Co.; 1969.

[10] Puntanen S, Styan GPH, Isotalo J. Matrix tricks for linear statistical models. Springer Berlin Heidelberg; 2011.

[11] Rao CR. Linear statistical inference and its applications. 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ, USA; 2002.

[12] Anderson TW. An introduction to multivariate statistical analysis. 3rd ed. John Wiley & Sons, Inc., Hobuken; 2003.

[13] Cook RD, Weisberg S. An introduction to regression graphics. John Wiley & Sons, Inc., New York, NY; 1994.

[14] Hocking RR. Methods and applications of linear models. Regression and the Analysis of Variance. 2nd ed. John Wiley & Sons, Inc., Hoboken, NJ, USA; 2003.

[15] Rencher AC, Schaalje GB. Linear models in statistics. 2nd ed. John Wiley & Sons, Hoboken, NJ, USA; 2008.

[16] Cochran WG. The omission or addition of an independent variate in multiple linear regression. J. R. Stat. Soc. Suppl. 1938; 5:171-176.

[17] Quenouille MH. An application of least squares to family diet surveys. Econometric. 1950; 18:27-44.

[18] Bingham NH, Fry JM. Regression, linear models in statistics. Springer, London; 2010.

[19] Chatterjee S, Hadi AS. Sensitivity analysis in linear regression. John Wiley & Sons, New York, NY; 1988.

[20] Seber GAF. Linear regression analysis. 2nd ed. New York: John Wiley & Sons; 2003.

[21] Dempster AP. Covariance selection. Biometrics. 1972;28:157-75.

[22] Allison RD. The impact of random predictors on comparisons of coefficients between models: Comment on clogg, petkova, and haritou. Am. J Sociol. 1995;100(5):1294-1305.

[23] Clogg CC, Petkova E, Cheng T. Reply to allison: More on comparing regression coefficients. Am J Sociol. 1995; 100(5):1305-1312.

[24] Clogg CC, Petkova E, Haritou A. Statistical methods for comparing regression coefficients between models. Am J Sociol. 1995; 100(5):1261-1293.

[25] Frank KA. Impact of a confounding variable on a regression coefficient. Sociol Methods & Research. 2000;29(2):147-194.

[26] Davier A. Large sample tests for comparing regression coefficients in models with normally distributed variables. Research Report RR-03-19, Princeton, NJ 08541, Educational Testing Service, Princeton, NJ; July 2003.

[27] Yan J, Aseltine RH, Harel O. Comparing regression coefficients between nested linear models for clustered data with generalized estimating equations. J Educat Behavioral Statistics. 2013;38(2):172-189.

[28] Bickel R. Multilevel analysis for applied research. Its Just Regression! Methodology in the Social Sciences. The Guilford Press, New York; 2007.

[29] Hox J. Multilevel analysis, techiques and applications, quantitaive methodology for businaess and management. Lawrence Erlbaum Accociates, Mahwah; 2002.

[30] Lee PH. Should we adjust for a confounder if empirical and theoretical criteria yield contradictory results? A Simulation Study. Sci Rep. 2014;4.
Article number: 6085.

[31] Maldonaldo G. Toward a clearer understanding of causal concepts in epidemiology. Ann Epidemiol. 2013;23:743-749.

[32] Anderson S, Auquier A, Hauck WW, Oakes D, Vandaele W, Weisberg HI. Statistical Methods for comparative studies. Techniques for Bias Reduction. John Wiley & Sons; 1980.

[33] Greenland S, Pearl J. Adjustments and their consequences-collapsibility analysis using graphical models. Int Stat Rev. 2011;79(3):401-426.

[34] Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci. 1999; 14(1):29-46.

[35] Pearl J. Why there is no statiistical test for confounding, why many think there is, and why they are almost right. Technical Report R-256; 1998.

[36] Steyer R, Gabler S, von Davier A, Nachtigall C. Causal regression models II: Unconfoundedness and causal unbiasedness. Methods of Psychological Research Online. 2000;5(3):55-87.

[37] Wakkee M, Hollestein LM, Nijsten T. Multivariable analysis. J Invest Dermatol. 2014;134(e20).

[38] Becker C, Fried R, Kuhnt S. Robustness and complex data structures: Festschrift in honour of ursula gather. springerlink : Bücher. Springer Berlin Heidelberg; 2014.

[39] Cox DR, Wermuth N. An approximation to maximum likelihood estimates in reduced models. Biometrika. 1990;77(4):747-761.

[40] Wermuth N, Cox DR. Joint response graphs and separation induced by triangular systems. J Royal Stat Soc B. 2004;66(3):687-717.

[41] Wermuth N, Marchetti GM, Cox DR. Triangular systems for symmetric binary variables. Electronic J Stat. 2009;3:932-955.

[42] Vellaisamy P, Vijay V. Collapsibility of regression coefficients and its extensions. J Statist Plann Inference, 2008;138(4):982-994.

[43] Hosman CA, Hansen BB, Holland PW. The sensitivity of linear regression coefficients confidence limits to the omission of a confounder. Ann Appl Stat. 2010;4(2):849-870.

[44] Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Lawrence Erlbaum Associates, Mahwah; 2003.

[45] Stuart A, Ord JK, Arnold S. Kendall's advanced theory of statistics: Volume 2A: Classical inference and the linear model. V. 2 in Kendall's library of statistics. Hodder Arnold; 1999.

[46] Marchetti GM, Wermuth N. Matrix representations and independencies in directed acyclic graphs. Ann Stat, apr 2009;37(2):961-978.

[47] Wermuth N. Probability distributions with summary graph structure. Bernoulli. 2011;17(3):845-879.

[48] Wermuth N, Cox DR, Marchetti GM. Covariance chains. Bernoulli. 2006;12(5):841-862.

[49] Panov VG, Varaksin AN. Relation between the coefficients of simple and multiple regression models. Siberian Math J. 2010;51(1):162-167.

[50] Affuso E, Caudill SP, Mixon Jr. FG. Omitted variables bias in regime-switching models with slope-constrained estimators: Evidence from monte carlo simulations. J Stat and Econometric Methods. 2013;2(3):49-55.

[51] Leighter JE, Inoue T. Solving the omitted variables problem of regression analysis using the relative vertical position of observations. Adv in Decision Sci; 2012. Article ID 728980.

[52] Suparman Y, Folmer H, Oud JHL. Hedonic price models with omitted variables and measurement errors: A constrained autoregressionstructural equation modeling approach with application to urban indonesia. J Geographical Sys. 2014;16(1):49-70.

[53] Ouellette DV. Schur complements and statistics. Linear Algebra and its Applications. 1981;36:187-295.

[54] Zhang F. ed. The schur complement and its applications. Numerical Methods and Algorithms. Springer New York, New York, NY; 2005.

[55] DeMaris A. Regression with social data. modelling continuous and limited response variables. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ; 2004.

[56] Draper NR, Smith H. Applied regression analysis. 3rd ed .John Wiley & Sons; 1998.

[57] Ehrenberg ASC. Data reduction: Analysing and interpreting statistical data. Wiley-Interscience publication. Wiley; 1975.

[58] Warner RM. Applied Statistics : From Bivariate through Multivariate Techniques. SAGE Publications, Los Angeles; 2008.

[59] Scherbo AP. Environment and health: Approaches to assessment of health risk. St-Petersburg MAPO, Saint-Petersburg; 2002.

[60] McNamee R. Regression modelling and other methods to control confounding. Occup Environ Med. 2005;62:500-506.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/11390