ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
20, 1 (1988), pp. 53–65

C. PLATT and Z. PAPRZYCKI (Olsztyn)

ON ESTIMATION OF MULTIPLE REGRESSION COEFFICIENTS BY THE p-POINT METHOD

1. Introduction. The classical approach to estimating the parameters of a regression line is via the least squares method. Another approach to this problem was presented by Hellwig [1]. The method proposed by him is called the *two-point method*. In this method, the observed values (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$ of the bivariate continuous random variable (X, Y) are divided into two sets I and II. The point (x_i, y_i) belongs to the set I if $x_i > \bar{x}$, where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

are the coordinates of the gravity center for the whole set of observations. Let $(\bar{x}_{\mathrm{I}}, \bar{y}_{\mathrm{I}})$ be the gravity center of the set I. In this case the expression

$$\hat{a}_{y/x} = \frac{\bar{y}_{\mathrm{I}} - \bar{y}}{\bar{x}_{\mathrm{I}} - \bar{x}}$$

is an estimator of the regression coefficient $a_{y/x}$. This estimator is unbiased and consistent but its efficiency is less than the efficiency of the least squares estimator. However, we find this estimator very easy to compute and that is why we deal with it. A similar problem was considered earlier by Wald [4]. Wald remarks that the division of a sample may be arbitrary, which leads us to a certain class of estimators. The purpose of our paper is to generalize the results of Hellwig ([1], [2]) and Wald [4] for the multiple regression case.

2. Determination of estimators of regression coefficients by the p-point method. We assume the following multiple regression model:

(1) $\quad Y_i = a_0 + a_1(x_{i1} - \bar{x}_1) + \ldots + a_p(x_{ip} - \bar{x}_p) + \varepsilon_i \quad$ for $i = 1, 2, \ldots, n$,

where $(Y_1, \ldots, Y_n)$ are the observed random variables, $a_0, \ldots, a_p$ are unknown fixed parameters, $x_{i1}, \ldots, x_{ip}$ are constants determined by the conditions of the i-th trial, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent, identically distributed random variables with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{D}^2(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$.

We define the sample as a set of the form

$$C_0 = \{(x_{i1}, \ldots, x_{ip}, y_i): i = 1, \ldots, n\},$$

where y_i is the observed value of Y_i. The set C_0 is divided in p manners into two disjoint subsets each time, i.e., if the pairs of subsets $C_1, C_1', \ldots, C_p, C_p'$ are the results of divisions, then for each pair C_k, C_k' we have

$$C_k \cup C_k' = C_0 \quad \text{and} \quad C_k \cap C_k' = \emptyset.$$

Let

$$\mathcal{C}_r = \{i \in \mathcal{C}_0: (x_{i1}, \ldots, x_{ip}, y_i) \in C_r\}, \qquad k_r = \operatorname{card} C_r,$$

where $\mathcal{C}_0 = \{1, \ldots, n\}$, and let

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{x}_s = \frac{1}{n}\sum_{i=1}^{n} x_{is}, \quad s = 1, 2, \ldots, p,$$

$$\bar{y}^r = \frac{1}{k_r}\sum_{i\in\mathcal{C}_r} y_i, \quad \bar{x}_s^r = \frac{1}{k_r}\sum_{i\in\mathcal{C}_r} x_{is}, \quad r, s = 1, \ldots, p.$$

The points $g_0 = (\bar{x}_1, \ldots, \bar{x}_p, \bar{y})$ and $g_r = (\bar{x}_1^r, \ldots, \bar{x}_p^r, \bar{y}^r)$ are called the *gravity centers* of the sets C_0 and C_r, respectively.

Let

$$y - \bar{y} = \hat{a}_1(x_1 - \bar{x}_1) + \ldots + \hat{a}_p(x_p - \bar{x}_p) \tag{2}$$

be the p-dimensional hyperplane passing through the points $g_0, g_1, \ldots, g_p$. In this case we set up the coefficients of (2) as the estimators of the regression coefficients in the model (1). If the matrix

$$W = \begin{bmatrix} \bar{x}_1^1 - \bar{x}_1 & \bar{x}_2^1 - \bar{x}_2 & \ldots & \bar{x}_p^1 - \bar{x}_p \\ \ldots & \ldots & \ldots & \ldots \\ \bar{x}_1^p - \bar{x}_1 & \bar{x}_2^p - \bar{x}_2 & \ldots & \bar{x}_p^p - \bar{x}_p \end{bmatrix}$$

is nonsingular, then the hyperplane (2) is determined uniquely. Therefore, we assume that the matrix W is nonsingular throughout this paper. Under the above assumption, the proposed estimators of the regression coefficients are

$$\hat{a}_0 = \bar{y}, \quad \hat{a}_i = W_i/W, \; i = 1, \ldots, p, \tag{3}$$

where $W \neq 0$ is the determinant of the matrix W, and W_i is the determinant obtained from W by replacing the i-th column with $(\bar{y}^1 - \bar{y}, \ldots, \bar{y}^p - \bar{y})^T$.

THEOREM 1. *The gravity centers g_r, g_0, g_r' of the sets C_r, C_0, C_r', respectively, lie along one straight line.*

The theorem follows from the equation

$$g_0 = \frac{k_r}{n} g_r + \left(1 - \frac{k_r}{n}\right) g'_r.$$

It is easy to see from Theorem 1 that it is sufficient to take into consideration one of the subsets C_k or C'_k for $k = 1, \ldots, p$. In this connection we denote the chosen subsets by $C_1, \ldots, C_p$, respectively. These subsets do not have to be disjoint.

The presented method of division determines a class of estimators because of the free-choice division of the set C_0.

3. Properties of the estimators obtained by the p-point method. Model (1) has a familiar form in the matrix notation:

(4)
$$y = Za_0 + \varepsilon,$$

where

$$y = (y_1, \ldots, y_n)^T, \qquad \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T,$$

$$a_0 = (a_0, a_1, \ldots, a_p)^T,$$

$$Z = \begin{bmatrix} 1 & x_{11} - \bar{x}_1 & \ldots & x_{1p} - \bar{x}_p \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} - \bar{x}_1 & \ldots & x_{np} - \bar{x}_p \end{bmatrix}.$$

According to the assumptions of model (1), the covariance matrix of the vector ε is of the form $\sigma^2 I$, where I is the identity matrix. The estimator

$$\hat{a}_0 = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_p)^T$$

of the vector of parameter a_0 is the linear function of the vector y, since

(5)
$$\hat{a}_0 = (W^*)^{-1} ry,$$

where

$$r = \begin{bmatrix} (1/n)\, 1 \\ (1/k_1)\, 1^1 - (1/n)\, 1 \\ \ldots \\ (1/k_p)\, 1^p - (1/n)\, 1 \end{bmatrix}, \qquad W^* = \begin{bmatrix} 1 & 0^T \\ 0 & W \end{bmatrix},$$

and 1 is a row vector of order n with all elements equal to unity. The row vector 1^l consists of k_l elements equal to unity and $n - k_l$ elements equal to zero. The r-th element of the vector 1^l is 1 if the r-th observation belongs to the set C_l and is 0 otherwise.

THEOREM 2. *The vector* $\hat{a}_0$ *is the unbiased estimator of the vector* a_0.

Proof. By virtue of (4) we have $\mathrm{E}(y) = Za_0$. Thus

$$\mathrm{E}(\hat{a}_0) = (W^*)^{-1} rZa_0 = (W^*)^{-1} W^* a_0 = a_0.$$

THEOREM 3. *The covariance matrix of the vector* $\hat{a}_0$ *is of the form*

$$\Sigma^* = \sigma^2 (W^*)^{-1} A^* [(W^*)^{-1}]^T,$$

where

$$A^* = \begin{bmatrix} 1/n & 0^T \\ 0 & A \end{bmatrix} = rr^T,$$

$$A = \left[\frac{k_{rs}}{k_r k_s} - \frac{1}{n}\right], \quad r, s = 1, \ldots, p,$$

$$k_{rs} = \begin{cases} \operatorname{card}(C_r \cap C_s) & \text{if } r \neq s, \\ k_r & \text{if } r = s. \end{cases}$$

From (5) we obtain

$$\mathrm{D}^2(\hat{a}_0) = (W^*)^{-1} r\mathrm{D}^2(y) r^T [(W^*)^{-1}]^T = \sigma^2 (W^*)^{-1} rr^T [(W^*)^{-1}]^T.$$

Theorem 3 implies that the estimator $\hat{a}_0 = \bar{y}$ is uncorrelated with the estimators $\hat{a}_i$ for $i = 1, \ldots, p$.

THEOREM 4. *Let* $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_p)^T$ *and* $a = (a_1, \ldots, a_p)^T$. *If*

(i) $$\lim_{n \to \infty} k_r/n = d_r, \quad \text{where } 0 < d_r < 1 \quad \text{for } r = 1, \ldots, p,$$

(ii) $$\liminf_{n \to \infty} |W| > 0,$$

(iii) $$|x_i^j| < M \quad \text{for some } M \text{ and } i, j = 1, \ldots, p,$$

then, for every vector $\lambda^T = (\lambda_1, \ldots, \lambda_p)$, $\lambda^T \hat{a}$ *converges stochastically towards* $\lambda^T a$.

Proof. Of course, $\mathrm{E}(\lambda^T \hat{a}) = \lambda^T a$ and

$$\begin{aligned} \mathrm{D}^2(\lambda^T \hat{a}) &= \mathrm{D}^2(\lambda_1 \hat{a}_1 + \ldots + \lambda_p \hat{a}_p) \\ &= \sum_{k=1}^{p} \sigma_{kk} \lambda_k^2 + \sum_{k=1}^{p} \sum_{l=1}^{p} \sigma_{kl} \lambda_k \lambda_l = \lambda^T \Sigma \lambda, \end{aligned}$$

where

$$\Sigma = [\sigma_{ij}] = \frac{\sigma^2}{W^2} [W_i^T A W_j] = \sigma^2 W^{-1} A (W^{-1})^T$$

is the covariance matrix of the vector $\hat{a}$. The matrix Σ is a generalized quadratic form. The diagonal elements of Σ are positive definite quadratic forms with the matrix A. The off diagonal elements of Σ are bilinear forms with the same matrix. Furthermore

$$\lim_{n\to\infty} a_{rr} = \lim_{n\to\infty}\left(\frac{1}{k_r}-\frac{1}{n}\right) = \lim_{n\to\infty}\frac{1}{k_r} = \lim_{n\to\infty}\frac{1}{n}\lim_{n\to\infty}\frac{n}{k_r} = 0,$$

$$\lim_{n\to\infty} a_{rs} = \lim_{n\to\infty}\left(\frac{k_{rs}}{k_r k_s}-\frac{1}{n}\right) = \lim_{n\to\infty}\frac{k_{rs}}{k_r k_s} \leqslant \lim_{n\to\infty}\frac{\max(k_r, k_s)}{k_r k_s} = 0.$$

From the above facts and from conditions (ii) and (iii) we obtain

$$\lim_{n\to\infty} \mathrm{D}^2(\lambda^T \hat{a}) = 0.$$

The assertion of Theorem 4 follows from Chebyshev's inequality.

From this theorem we infer immediately that the estimator $\hat{a}$ is consistent. Note that the estimator $\hat{a}_0 = \bar{y}$, as a sample mean, is also consistent. Let us represent the set C_0 as a sum of disjoint subsets $B_1, \ldots, B_m$, where each B_j is of the form

$$C_1^{j_1} \cap C_2^{j_2} \cap \ldots \cap C_p^{j_p},$$

where $C_k^{j_k}$ means C_k or C_k'. Then $m = 2^p$.

THEOREM 5. *If the assumptions of Theorem 4 hold and if*

(i) $\lim\limits_{n\to\infty} \bar{x}_r^c$ *exists,*

(ii) *if* $l_k = \operatorname{card} B_k$, *then*

$$\lim_{n\to\infty} l_k/n = c_k, \quad \textit{where } 0 < c_k < 1,\ k = 1, \ldots, 2^p,$$

(iii) $\lim\limits_{n\to\infty} nA = A^c$, *where* A^c *is a positive definite matrix,*

then the joint distribution of the random vector

$$(T_1^{(n)}, \ldots, T_p^{(n)}),$$

where

$$T_i^{(n)} = \frac{\hat{a}_i^{(n)} - a_i}{\sqrt{\mathrm{D}^2(\hat{a}_i^{(n)})}} \tag{6}$$

is asymptotically normal $N(0, \Sigma_n)$ *with*

$$\Sigma_n = \left[\frac{W_r^T A W_s}{\sqrt{W_r^T A W_r}\sqrt{W_s^T A W_s}}\right].$$

In the proof of Theorem 5 we use the following central limit theorem for multivariate random variables from the monograph by Rao ([3], p. 165):

Let $X_1, X_2, \ldots$ be a sequence of independent random vectors of order k with $\mathrm{E}(X_i) = 0$ and $\operatorname{cov}(X_i) = \Sigma_i^n$, $i = 1, 2, \ldots$ If

$$\frac{1}{n}\sum_{i=1}^{n} \Sigma_i^n \to \Sigma \neq 0 \quad \text{as } n \to \infty \tag{7}$$

and, for every positive number τ,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\|X_i\|^2\colon \|X_i\| > \tau\sqrt{n}) = \frac{1}{n}\sum_{i=1}^{n} \int_{\|X_i\| > \tau\sqrt{n}} \|X_i\|^2\, dF_i \to 0, \tag{8}$$

where F_i is a distribution and $\|X_i\|$ is the Euclidean norm of the random vector X_i, then the sequence of distributions of the random vector $(X_1 + \ldots + X_n)/\sqrt{n}$ converges to the k-dimensional normal distribution $N(0, \Sigma)$.

Proof of Theorem 5. Note that

$$\begin{aligned}
\hat{a}_i - a_i &= \frac{1}{W}\sum_{r=1}^{n} (\bar{\varepsilon}^r - \bar{\varepsilon})\, W_{ri} \\
&= \frac{1}{W}\sum_{r=1}^{n} \left(\frac{1}{k_r}\sum_{j\in C_r} \varepsilon_j - \frac{1}{n}\sum_{j\in C_0} \varepsilon_j\right) W_{ri} \\
&= \frac{1}{W}\left[\sum_{j\in B_1} \varepsilon_j \left(\frac{W_{1i}}{k_1} + \ldots + \frac{W_{pi}}{k_p} - \frac{W_{1i}}{n} - \ldots - \frac{W_{pi}}{n}\right) + \ldots \right. \\
&\qquad \left. + \sum_{j\in B_m} \varepsilon_j \left(-\frac{W_{1i}}{n} - \ldots - \frac{W_{pi}}{n}\right)\right] \\
&= \frac{1}{W}\sum_{j=1}^{n} \varepsilon_j \alpha_{jn}^{(i)},
\end{aligned} \tag{9}$$

where $\alpha_{jn}^{(i)}$ is the expression in the brackets at ε_j. Substituting the last expression into formula (6) we can write $T_i^{(n)}$ as

$$T_i^{(n)} = \sum_{j\in B_1} \varepsilon_j \beta_{jn}^{(i)} + \ldots + \sum_{j\in B_m} \varepsilon_j \beta_{jn}^{(i)} = \sum_{j=1}^{n} \varepsilon_j \beta_{jn}^{(i)}, \tag{10}$$

where $\beta_{jn}^{(i)}$ are of the form

$$\beta_{jn}^{(i)} = \frac{\alpha_{jn}^{(i)}}{W\sqrt{\mathrm{D}^2(\hat{a}_i^{(n)})}}. \tag{11}$$

Since, for every $j \in B_r$, $\beta_{jn}^{(i)}$ have the same value, say $\beta_{rn}^{(i)}$, we obtain

$$\mathrm{D}^2(T_i^{(n)}) = \sigma^2 l_1 (\beta_{1n}^{(i)})^2 + \ldots + \sigma^2 l_m (\beta_{mn}^{(i)})^2,$$

where l_k for $k = 1, \ldots, 2^p$ is the cardinal number of the set B_k. Further, without loss of generality, we take into consideration only the non-zero elements l_k and $\beta_{kn}^{(i)}$. Thus

$$l_1 (\beta_{1n}^{(i)})^2 + \ldots + l_m (\beta_{mn}^{(i)})^2 = 1/\sigma^2,$$

and this yields

$$0 < l_k (\beta_{kn}^{(i)})^2 \leqslant 1/\sigma^2 \quad \text{for } k = 1, \ldots, n.$$

Hence

$$0 < \frac{l_k}{n} (\beta_{kn}^{(i)})^2 \leqslant \frac{1}{n\sigma^2}.$$

Since $0 < c_k < 1$, we get

$$\lim_{n\to\infty} (\beta_{kn}^{(i)})^2 = 0.$$

For $j = 1, \ldots, n$ we define the following vectors of order p:

(12) $$X_j^{(n)} = (\varepsilon_j \beta_{jn}^{(1)}, \ldots, \varepsilon_j \beta_{jn}^{(p)})^T \sqrt{n}.$$

Since, by assumption, ε_j for $j = 1, \ldots, n$ are independent random variables, the vectors $X_j^{(n)}$ are also stochastically independent. By (10) we have

$$(T_1^{(n)}, \ldots, T_p^{(n)})^T = (X_1^{(n)} + \ldots + X_n^{(n)})/\sqrt{n} = \Big(\sum_{j=1}^{n} \varepsilon_j \beta_{jn}^{(1)}, \ldots, \sum_{j=1}^{n} \varepsilon_j \beta_{jn}^{(p)}\Big)^T.$$

The definition of the vectors (12) and the assumption $\mathrm{E}(\varepsilon_j) = 0$ imply $\mathrm{E}(X_j^{(n)}) = 0$. Now we must prove that the vectors (12) satisfy conditions (7) and (8).

To prove (7) let us note that the elements of the covariance matrix Σ_j^n of the vector $X_j^{(n)}$ are

$$\mathrm{cov}(\varepsilon_j \beta_{jn}^{(i)} \sqrt{n}, \varepsilon_j \beta_{jn}^{(k)} \sqrt{n}) = n\sigma^2 \beta_{jn}^{(i)} \beta_{jn}^{(k)}.$$

Hence, by (9) and (11), the elements of the matrix $n^{-1} \sum_{j=1}^{n} \Sigma_j^n$ can be written as

$$\sigma^2 \sum_{j=1}^{n} \beta_{jn}^{(i)} \beta_{jn}^{(k)} = \sigma^2 \sum_{j=1}^{n} \frac{\alpha_{jn}^{(i)} \alpha_{jn}^{(k)}}{W^2 \sqrt{\mathrm{D}^2(\hat{a}_i^{(n)})} \sqrt{\mathrm{D}^2(\hat{a}_k^{(n)})}}$$

$$= \frac{\mathrm{cov}(\hat{a}_i^{(n)}, \hat{a}_k^{(n)})}{\sqrt{\mathrm{D}^2(\hat{a}_i^{(n)})} \sqrt{\mathrm{D}^2(\hat{a}_k^{(n)})}} = \frac{W_i^T A W_k^T}{\sqrt{W_i^T A W_i} \sqrt{W_k^T A W_k}} = \mathrm{cov}(T_i^{(n)}, T_k^{(n)}).$$

By the assumption of the theorem and using the above results we get

$$\lim_{n\to\infty} nA = A^c = \left[\frac{c_{rs}}{c_r c_s} - 1\right] \quad \text{for } r, s = 1, \ldots, p$$

and

(13) $$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \Sigma_j^n$$

$$= \lim_{n\to\infty} \frac{W_i^T A W_k}{\sqrt{W_i^T A W_i} \sqrt{W_k^T A W_k}} = \frac{W_{i0}^T A^c W_{k0}}{\sqrt{W_{i0}^T A^c W_{i0}} \sqrt{W_{k0}^T A W_{k0}}},$$

where W_{i0} is the limit vector of W_i. From the assumptions of the theorem it follows that not all elements of the vector W_{i0} are equal to zero and the denominators in the limit (13) are different from zero. Hence the proof of (7) is complete.

To prove (8), let τ be any positive number. Then

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(\|X_j^{(n)}\|^2\colon\ \|X_j^{(n)}\| > \tau\sqrt{n})\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}(n\varepsilon_j^2\|\beta_{jn}\|^2\colon\ |\varepsilon_j|\,\|\beta_{jn}\| > \tau)\\
&= \lim_{n\to\infty}\Big[\sum_{j\in B_1}\|\beta_{1n}\|^2\,\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{1n}\|}\Big)+\ldots\\
&\quad + \sum_{j\in B_m}\|\beta_{mn}\|^2\,\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{mn}\|}\Big)\Big]\\
&= \lim_{n\to\infty} l_1\|\beta_{1n}\|^2\,\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{1n}\|}\Big)+\ldots\\
&\quad + \lim_{n\to\infty} l_m\|\beta_{mn}\|^2\,\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{mn}\|}\Big)\\
&\leqslant \frac{p}{\sigma^2}\Big[\lim_{n\to\infty}\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{1n}\|}\Big)+\ldots+\lim_{n\to\infty}\mathrm{E}\Big(\varepsilon_j^2\colon\ |\varepsilon_j| > \frac{\tau}{\|\beta_{mn}\|}\Big)\Big],
\end{aligned}$$

where $\|\beta_{jn}\| = \sqrt{(\beta_{jn}^{(1)})^2+\ldots+(\beta_{jn}^{(p)})^2}$. This completes the proof since $\lim_{n\to\infty}(\beta_{jn}^{(i)})^2 = 0$ implies that each of the last $m = 2^p$ limits is equal to zero, i.e., (8) is true.

Remark. The assumption relating to the convergence of $\bar{x}_j^i$ as $n\to\infty$ is unnecessary for the convergence of the marginal distributions $T_i^{(n)}$ to the normal distribution $N(0, 1)$. In this case, the assumptions of Theorem 3 are sufficient.

THEOREM 6. *Let SS_p denote the residual error sum of squares if the method of estimation described in our paper is used and let SS_k denote the same in the least squares method. Then*

$$\lim_{n\to\infty}\frac{\mathrm{E}(SS_p)}{\mathrm{E}(SS_k)} = 1.$$

Proof. It is easy to verify that

$$\mathrm{E}(SS_p) = \sigma^2\Big(n-1+\frac{1}{W^2}\sum_{i=1}^{n}\sum_{r,s=1}^{p}(x_{is}-\bar{x}_s)(x_{ir}-\bar{x}_r)\,W_r^T A W_s\Big).$$

Furthermore, it is known that

$$E(SS_k) = \sigma^2(n-p-1).$$

These results imply the assertion of Theorem 6.

4. Applications.

EXAMPLE 1. In experimental designs in which the matrix of data has mutually orthogonal columns, the regression coefficients are mutually independent and their simple form assists the simplicity of computation. In particular, let us consider a p-factorial experiment of type 2^p in which the variables $x_1, \ldots, x_p$ have two values only, 1 or -1, and all treatments occur the same number of times. Using divisions of the observed values according to the formula

$$C_k = \{(x_1, \ldots, x_p, y): x_k = -1\},$$

we obtain mutually independent regression coefficients of the form

$$\hat{a}_k = \frac{\bar{y}^k - \bar{y}}{\bar{x}_k^k} \quad \text{for } k = 1, \ldots, p.$$

In this case, these coefficients and the regression coefficients obtained by the least squares method are identical.

EXAMPLE 2. In a map, a square with side length equal to 90 m was fixed. Then this region was divided into smaller squares with side lengths equal to 10 m. Subsequently, a rectangular coordinate system, in which the axes OX_1 and OX_2 divide the opposite sides of the initial square into two, was constituted. For every vertex of the squares we are interested in the following values: y_i — height in relation to the sea level, x_{i1}, x_{i2} — coordinates in the system set up. Those quantities are given in Table 1. With the regression plane we want to describe the topographical surface of the fixed region. For this purpose we use the method proposed in this paper. Let C_0 be the set of those points the coordinates of which are given in Table 1. As a result of divisions of the set C_0 we obtain two sets

$$C_1 = \{(x_{i1}, x_{i2}, y_i): x_{i1} < 0\}, \qquad C_2 = \{(x_{i1}, x_{i2}, y_i): x_{i2} < 0\}$$

with the same cardinal number $k_1 = k_2 = 50$.

The symmetric position of the points describing the topographical surface in relation to the system $OX_1 OX_2$ implies that the sums of numbers in columns x_{i1} and x_{i2} are equal to zero. Hence $\bar{x}_1 = 0$, $\bar{x}_2 = 0$ and $\bar{x}_1^1 = -25$, $\bar{x}_2^1 = 0$, $\bar{x}_2^2 = -25$, $\bar{x}_1^2 = 0$. The quantities $\bar{x}_1^1$ and $\bar{x}_2^2$ depend on the number of points describing the surface as follows:

$$\bar{x}_1^1 = \bar{x}_2^2 = -\frac{a\sqrt{n}}{4(\sqrt{n}-1)},$$

TABLE 1. The coordinates of the vertices of squares and their heights in relation to the sea level of a map

x_{i1}	x_{i2}	y_i	C_1	C_2	x_{i1}	x_{i2}	y_i	C_1	C_2
−45	+45	110.3	+		5	45	108.1		
−45	+35	110.7	−		5	35	108.7		
−45	+25	110.0	+		5	25	109.5		
−45	+15	111.3	+		5	15	110.5		
−45	5	111.6	+		5	5	111.7		
−45	−5	112.5	+	+	5	−5	112.8		+
−45	−15	113.2	+	+	5	−15	114.0		+
−45	−25	113.4	+	+	5	−25	114.9		+
−45	−35	113.6	+	+	5	−35	115.4		+
−45	−45	113.7	+	+	5	−45	115.7		+
−35	45	109.3	+		15	45	107.9		
−35	35	109.8	+		15	35	108.6		
−35	25	110.2	+		15	25	109.5		
−35	15	110.7	+		15	15	110.6		
−35	5	111.5	+		15	5	111.8		
−35	−5	112.4	+	+	15	−5	113.1		
−35	−15	113.4	+	+	15	−15	114.3		+
−35	−25	113.7	+	+	15	−25	115.1		+
−35	−35	114.0	+	+	15	−35	115.5		+
−35	−45	114.0	+	+	15	−45	115.9		+
−25	45	108.8	+		25	45	107.7		
−25	35	109.3	+		25	35	108.5		
−25	25	109.8	+		25	25	109.5		
−25	15	110.5	+		25	15	110.7		
−25	5	111.5	+		25	5	112.0		
−25	−5	112.5	+	+	25	−5	113.3		+
−25	−15	113.6	+	+	25	−15	114.6		+
−25	−25	114.2	+	+	25	−25	115.2		+
−25	−35	114.4	+	+	25	−35	115.6		+
−25	−45	114.5	+	+	25	−45	116.1		+
−15	45	108.4	+		35	45	107.4		
−15	35	108.9	+		35	35	108.4		
−15	25	109.6	+		35	25	109.4		
−15	15	110.5	+		35	15	110.7		
−15	5	111.5	+		35	5	112.1		
−15	−5	112.5	+	+	35	−5	113.6		+
−15	−15	113.7	+	+	35	−15	114.7		+
−15	−25	114.5	+	+	35	−25	115.3		+
−15	−35	115.1	+	+	35	−35	115.7		+
−15	−45	115.1	+	+	35	−45	116.2		+
−5	45	108.3	+		45	45	106.9		
−5	35	108.8	+		45	35	108.3		
−5	25	109.5	+		45	25	109.4		
−5	15	110.5	+		45	15	110.8		
−5	5	111.6	+		45	5	112.3		
−5	−5	112.6	+	+	45	−5	113.8		+
−5	−15	113.8	+	+	45	−15	114.8		+
−5	−25	114.7	+	+	45	−25	115.3		+
−5	−35	115.2	+	+	45	−35	115.7		+
−5	−45	115.5	+	+	45	−45	116.2		+

where n is the number of points, and a is the side length of the square containing the described topographical region. Furthermore, all the assumptions of the paper are fulfilled:

1. $\lim\limits_{n\to\infty} k_1/n = \lim\limits_{n\to\infty} k_2/n = 1/2$, $\lim\limits_{n\to\infty} k_{12}/n = 1/4$;
2. $\liminf\limits_{n\to\infty} |W| = a^2/16 > 0$;
3. there exists $\lim\limits_{n\to\infty} \bar{x}^i_j$ for $i, j = 1, 2$, and

$$\lim_{n\to\infty} \bar{x}^1_1 = \lim_{n\to\infty} \bar{x}^2_2 = \lim_{n\to\infty} \frac{-a\sqrt{n}}{4(\sqrt{n}-1)} = -\frac{a}{4},$$

$$\lim_{n\to\infty} \bar{x}^1_2 = \lim_{n\to\infty} \bar{x}^2_1 = 0;$$

4. $W \neq 0$, where

$$W = \begin{vmatrix} \bar{x}^1_1 & 0 \\ 0 & \bar{x}^2_2 \end{vmatrix} = \bar{x}^1_1 \bar{x}^2_2 .$$

Thus

$$\hat{a}_1 = \frac{\begin{vmatrix} \bar{y}^1 - \bar{y} & 0 \\ \bar{y}^2 - \bar{y} & \bar{x}^2_2 \end{vmatrix}}{\bar{x}^1_1 \bar{x}^2_2} = \frac{\bar{y}^1 - \bar{y}}{\bar{x}^1_1},$$

$$\hat{a}_2 = \frac{\begin{vmatrix} \bar{x}^1_1 & \bar{y}^1 - \bar{y} \\ 0 & \bar{y}^2 - \bar{y} \end{vmatrix}}{\bar{x}^1_1 \bar{x}^2_2} = \frac{\bar{y}^2 - \bar{y}}{\bar{x}^2_2}.$$

In Table 1 the sign + in column C_i, $i = 1, 2$, means that the point belongs to the set C_i. Since the values $\bar{x}_i$ are determined by the choice of the coordinate system and by the division of the set of points, it remains to calculate only $\bar{y}$, $\bar{y}^1$, $\bar{y}^2$, which are equal to

$$\bar{y} = 112.135, \quad \bar{y}^1 = 111.994, \quad \bar{y}^2 = 114.372,$$

respectively, and

$$a_0 = 112.135, \quad a_1 = 0.0056, \quad a_2 = -0.0895.$$

Hence we have the following regression equation:

$$y = 112.135 + 0.0056\, x_1 - 0.0895\, x_2 .$$

For comparison, the regression equation obtained by the least squares method is of the form

$$y = 112.135 + 0.0049\, x_1 - 0.0852\, x_2 .$$

The residual error sums of squares have values

$$SS_p = 51.043 \quad \text{and} \quad SS_k = 49.468,$$

respectively. The rate $SS_p/SS_k = 1.0318$ is close to unity.

In conclusion we want to make some essential remarks.

The estimators of the regression coefficients discussed in the present paper are less efficient than the estimators obtained by the least squares method, but in the large sample case the efficiency of these estimators is sufficiently good. As was noted above, this efficiency depends on the division of the set C_0 and as a measure of this efficiency we can take the determinant of the covariance matrix

$$|\Sigma| = \sigma^2 \frac{|A|}{W^2}.$$

Since finding the minimum value of this determinant with respect to the numbers of elements of the sets C_i is very complicated, nine simulated samples have been examined. In each case, $\det \Sigma$ attained the minimum value when the set C_0 was divided into two subsets with equal or approximately equal numbers of elements. In the $(p+1)$-dimension regression case, where the sample has the form of the set $C_0 \subset R^{p+1}$, this division can be obtained by means of the regression established for the set $\{(x_{i1}, \ldots, x_{ip}): i = 1, \ldots, n\} \subset R^p$. The construction of these divisions, when $p > 0$ and $x_{i1}, \ldots, x_{ip}$ are arbitrary, may be rather complicated from the numerical point of view, however in many cases the divisions are determined by the conditions of the experimental design. For instance, this situation arises in the factorial design of the experiment discussed in Section 4. In both of the above cases, the assumptions of the model (1) are fulfilled because we assume that $x_{i1}, \ldots, x_{ip}$ are constants determined by the experimental designs. Also the matrices of the experiments are orthogonal, which involves the independence of the estimators of the regression coefficients.

It is to be emphasized that, in the situation when the ratio p/n is close to unity, the division of the set C_0 has a great effect on the values of the estimators and the efficiency of these estimators is very small. The method presented in our paper is convenient for using when n is large and p/n is close to zero.

It is also of interest to note that the good robustness is a favourable property of these estimators. The authors consider this problem in a separate paper.

References

[1] Z. Hellwig, *Wyznaczanie parametrów regresji liniowej metodą dwóch punktów*, Zastos. Mat. 3 (1956), pp. 66–81.
[2] — *Aproksymacja stochastyczna*, PWE, Warszawa 1965.
[3] C. R. Rao, *Modele liniowe statystyki matematycznej*, PWN, Warszawa 1982.
[4] A. Wald, *The fitting of straight lines if both variables are subject to error*, Ann. Math. Statist. 11 (1940), pp. 284–300.

DEPT. OF APPLICATION OF MATHEMATICS
AND MATHEMATICAL STATISTICS
AGRICULTURAL AND TECHNICAL ACADEMY
10-740 OLSZTYN

Received on 1983.06.22;
revised version on 1986.03.13
