Listy Biometryczne - Biometrical Letters
Vol. 36(1999), No. 1, 31-36

The optimal plans in determination of fractions

Radosław Kala and Izabela Arcisz

Department of Mathematical and Statistical Methods,
Agricultural University of Poznań
Wojska Polskiego 28, 60-637 Poznań, Poland

SUMMARY

Chemical elements, like heavy metals, which occur in soils as a result of environment contamination, take different forms, called fractions. To determine the structure of contamination, a sequential chemical extraction is applied. In that way the researcher can determine the total content of the searched element as well the contents of all its particular fractions. Since the whole process is expensive and time consuming, the total number of measurements is confined. Under these assumptions, we indicate the optimal plans of determination of fractions with regard to the precision of the results obtained.

KEY WORDS: A-optimality, D-optimality, best linear unbiased estimator.

1. Introduction

Investigating the soils contamination by various metals, e.g. lead, copper, zinc, etc., the contents of these chemical elements are evaluated. Since each such element in soil occur in many different forms, called fractions, the determination of the total content of the searched element is usually supplemented by determination, from a parallel soil sample, of the contents of its particular fractions. Often, the rest of the chemical element is also measured as the residual fraction. To improve the precision, each measurement is usually repeated many times.

The traditional way of analyzing such data consists in calculating means of repeated measurements, which supply the estimates of the total content as well the contents of all fractions. In general, however, the sum of estimated fractions, together with the residual one, is not equal to the estimate of the total content. As it can be seen from the data presented by McLaren and Crawford (1973), the differences, between the sum of fractions and the total, exceed 10 per cent of the total content.

A different solution of this problem was proposed by Kala and Arcisz (1996). They incorporated into estimation procedure the condition of balance between the total content and the sum of all its fractions, which have improved precision of the determinations. Nevertheless, it is of some interest to maximize the efficiency of the estimates by proper planning of the numbers of replicates. The aim of the paper is to characterize A- and D-optimal plans for such investigation under the assumption that all fractions, together with the residual one, are determined.

2. Model and estimation

Let us assume that for the element being under investigation, we can distinguish $k, k > 1$, its different fractions, together with the residual one. Let y_{0j} denote the j-th measurement of the total content, where $j = 1, 2, ..., n_0$, and let y_{lj} denote j-th measurement of the l-th fraction, where $j = 1, 2, ..., n_l$, $l = 1, 2, ..., k$. Moreover, let us assume that the analytic methods of determining the total content and its particular fractions are conducted independently, with the same error variance σ^2.

All measurements, $N = \sum_{l=0}^{k} n_l$ in number, can be included in the following model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

in which $\mathbf{y} = (y_{01}, y_{02}, ..., y_{kn_k})'$ is a vector of all measurements, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)'$ is a vector of unknown contents of fractions, e is a vector of independent random errors, with zero expectation and the same variance σ^2, and $\mathbf{X}$ is an $N \times k$ design matrix of the form

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{1}$$

The first n_0 rows of $\mathbf{X}$ result from the fact that the total content is the sum of all its fractions. Moreover, observe that $\mathbf{X}$ is of full column rank, which ensures estimability of any linear function of parameters.

The best linear unbiased estimator of β was established by Kala and Arcisz (1996). For the content of the l-th fraction it can be expressed in the form

$$\hat{\beta}_l = \bar{y}_l + \frac{1}{n_l \tilde{n}_0} \left(\bar{y}_0 - \sum_{i=1}^{k} \bar{y}_i \right), \tag{2}$$

where $\tilde{n}_0 = \sum_{i=0}^{k} 1/n_i$ and $\bar{y}_l = \sum_{i=1}^{n_l} y_{li}/n_l$ is the mean of measurements for the l-th fraction, while $\bar{y}_0 = \sum_{i=1}^{n_0} y_{0i}/n_0$ is the mean for the total content. The variance of $\hat{\beta}_l$ has the form

$$\operatorname{var}(\hat{\beta}_l) = \frac{\sigma^2}{n_l} \left(1 - \frac{1}{n_l \tilde{n}_0} \right). \tag{3}$$

It is easy to note that the variance (3) is smaller than the variance of the mean $\bar{y}_l$, $\operatorname{var}(\bar{y}_l) = \sigma^2/n_l$. Moreover, since all fractions are observed, the total content can be estimated by the sum $\sum_{l=1}^{k} \hat{\beta}_l$. This estimator has the variance $\sigma^2 \tilde{n}/(n_0 \tilde{n}_0)$, where $\tilde{n} = \sum_{i=1}^{k} 1/n_i$. Since $\tilde{n}/\tilde{n}_0 < 1$, the variance $\sigma^2 \tilde{n}/(n_0 \tilde{n}_0)$ is also smaller than the variance of the mean $\bar{y}_0$.

3. A-optimal plans

The methods of extracting fractions are expensive and time consuming. This is the main cause of confining the total number N of measurements. However, given N, it is possible to propose different plans of replications $\{n_0, n_1, ..., n_k\}$. Thus arises the problem of choosing the optimal plan. One solution leads to A-optimal plans, which ensure that the average of variances of estimators $\hat{\beta}_l$, $l = 1, 2, ..., k$, will be minimized (cf. Pukelsheim, 1993, p.137). Since k is fixed, the searched plan minimizes the sum of variances of the form (3), i.e. the sum

$$\varphi = \sigma^2 \sum_{l=1}^{k} \frac{1}{n_l} \left(1 - \frac{1}{n_l \tilde{n}_0} \right).$$

The solution is given in the following

THEOREM 1. *The plan* $\{n_0, n_1, ..., n_k\}$, $n_l \geq 1$, $k > 1$, $N = \sum_{l=0}^{k} n_l$, *is A-optimal, if the measurements of all fractions are equireplicated, i.e.* $n_1 = n_2 = ... = n_k = n$, *while the number of replications of the total content* n_0 *fulfills the condition* $n = n_0 \left(1 + \sqrt{1+k}\right)$.

Proof. Using the Lagrange method, the problem reduces to minimization of the function

$$\Phi = \sum_{l=1}^{k} \frac{1}{n_l}\left(1 - \frac{1}{n_l \tilde{n}_0}\right) + \lambda\left(\sum_{l=0}^{k} n_l - N\right), \tag{4}$$

where λ is a multiplier. To simplify the function, let us introduce new variables

$$1/n_l = x_l, \quad l = 0, 1, ..., k. \tag{5}$$

Then the function (4) takes the form

$$\Phi = \sum_{l=1}^{k} x_l - \frac{1}{A}\sum_{l=1}^{k} x_l^2 + \lambda\left(\sum_{l=0}^{k} \frac{1}{x_l} - N\right),$$

where $A = \sum_{l=0}^{k} x_l$. Equating to zero the partial derivatives of Φ with respect to $x_1, x_2, ..., x_k$ and x_0, respectively, leads to the stationary equations

$$\begin{aligned} 1 + \frac{1}{A^2}\sum_{l=1}^{k} x_l^2 - \frac{2x_j}{A} &= \frac{\lambda}{x_j^2}, \quad j = 1, 2, ..., k \\ \lambda &= \frac{x_0^2}{A^2}\sum_{l=1}^{k} x_l^2, \end{aligned} \tag{6}$$

with the condition $N = \sum_{l=0}^{k} 1/x_l$. Comparing any two equations from the first set, implies, in particular, that $x_1 = x_2 = ... = x_k (= x)$. Thus the system (6) reduces to the form

$$1 + \frac{kx^2}{A^2} - \frac{2x}{A} = \frac{\lambda}{x^2}, \quad \lambda = \frac{x_0^2 k x^2}{A^2},$$

where now $A = x_0 + kx$ and $N = 1/x_0 + k/x$. Eliminating λ and A from the first equation, gives the equality

$$k + 2(x_0/x) = (x_0/x)^2,$$

which reduces to the condition $x_0 = \left(1 + \sqrt{1+k}\right) x$. Thus, in view of (5), the proof is completed. □

4. D-optimal plans

Another solution of the problem considered here is provided by D-optimality criterion. As it is well known (cf. Pukelsheim, 1993, p. 136), such a plan minimizes the generalized variance of the optimal estimator of $\boldsymbol{\beta}$. This generalized variance is equal

here to the determinant of $(\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X}$ is given in (1). The optimal plans are characterized in the following

THEOREM 1. *The plan* $\{n_0, n_1, ..., n_k\}, n_l \geq 1$, $k > 1$, $N = \sum_{l=0}^{k} n_l$, *is D-optimal, if all measurements are equireplicated, i.e.* $n_0 = n_1 = ... = n_k$.

Proof. It can easily be shown that the matrix $\mathbf{X}'\mathbf{X}$ takes here the form

$$\mathbf{X}'\mathbf{X} = \operatorname{diag}(n_1, n_2, ..., n_k) + n_0 \mathbf{1}_k \mathbf{1}_k'.$$

Using now the well known properties (see e.g. Mardia et. al., 1979, A.2.3 h, A.2.3 m), the determinant of $(\mathbf{X}'\mathbf{X})^{-1}$ can be expressed as

$$\left|(\mathbf{X}'\mathbf{X})^{-1}\right| = \frac{1}{n_0 \tilde{n}_0} \prod_{l=1}^{k} \frac{1}{n_l}.$$

Substituting new variables, as in the proof of Theorem 1, and introducing the Lagrange multiplier, the problem reduces to minimization of

$$\psi = \frac{x_0}{A} \prod_{l=1}^{k} x_l + \lambda \left(\sum_{l=0}^{k} \frac{1}{x_l} - N \right),$$

where $A = \sum_{l=0}^{k} x_l$. Comparing to zero the partial derivatives of ψ, we obtain the equations

$$\frac{A - x_j}{A^2 x_j} \prod_{l=1}^{k} x_l = \frac{\lambda}{x_j^2}, \quad j = 1, 2, ..., k$$

$$\lambda = \frac{x_0^2}{A^2} \left(\prod_{l=1}^{k} x_l \right) \left(\sum_{l=0}^{k} x_l \right), \tag{7}$$

with the condition $N = \sum_{l=0}^{k} 1/x_l$. As in the proof of Theorem 1, the first k equations imply that $x_1 = x_2 = ... = x_k (= x)$. Hence (7) reduces to the form

$$\frac{(A - x) x_0 x^k}{A^2 x} = \frac{\lambda}{x^2}, \quad \lambda = \frac{x_0^2 x^{k+1} k}{A^2},$$

where now $A = x_0 + kx$. Substituting λ from the second equation to the first one and eliminating A, implies $x_0 = x$, which completes the proof. □

5. Concluding remarks

The results obtained are in agreement with the intuition following directly from the definition of A- and D-optimality. In the case of D-optimality, when generalized variance is of main interest, all parametric functions are treated equally, which results in equal replications of measurements of the total content, as well as all its fractions. In the case of A-optimality, when the average of variances of separate estimates is preferred, each fraction must be equally replicated, while the measurement of the total content must be replicated much less. The precise condition is given in Theorem 1.

Of course, the established criteria are not applicable for all pairs (N, k). In the case of D-optimality it is sufficient that N is divisible by $k+1$. For A-optimality, the condition is more complicated. Since it is expressed by the square root of $k+1$, the exact solutions exist only for some k. Anyway, even if the exact solution does not exist, the established results indicate which plans can be better under the A- or D-optimality criterion.

References

Kala, R. and Arcisz, I. (1996). On fraction determination analysis. *XXVI Colloquium Biometryczne* **26**, 90-96. (in Polish).

Mardia, K.V., Kent, I.T. and Bibby, I.M. (1979). *Multivariate Analysis.* Academic Press, London.

McLaren, K.C. and Crawford, D.V. (1973). Studies on soil copper. The fractionation of copper in soils. *Journal of Soil Science* **24**, 172-181.

Pukelsheim, F. (1996). *Optimal Design of Experiments.* Wiley, New York.

Received 6 October 1998

Optymalne plany oznaczeń frakcji

Streszczenie

Pierwiastki występujące w glebie na skutek zanieczyszczenia środowiska, najczęściej metale ciężkie, przyjmują różne formy zwane frakcjami. Chcąc określić strukturę zanieczyszczenia prowadzi się sekwencyjne badania analityczne mające na celu oznaczenie ogólnej zawartości badanego pierwiastka jak również zawartości poszczególnych jego frakcji. Ponieważ badania takie są kosztowne i czasochłonne, całkowita liczba oznaczeń jest ograniczona. Przy tych założeniach wskazujemy optymalne plany oznaczeń frakcji z uwzględnieniem precyzji uzyskanych wyników.

Słowa kluczowe: A-optymalność, D-optymalność, najlepszy liniowy estymator nieobciążony.