ON THE ESTIMATION OF FUNCTIONAL RELATIONSHIPS FOR A FINITE UNIVERSE

J. C. Koop, Research Triangle Institute, N. C.

1. Introduction

In the modern theory of least squares, and the theories from which it evolved [12], the notion of residuals or the somewhat kindred notion of errors estimation, both of which have a strong empirical basis, play a basic role [2, 3]. Either one of these notions, in their relevent context, with the twin principles of unbiasedness and minimum variance, yield the solutions to the problems of fitting curves or surfaces, or adjustment of observations.

In this connection, in the theory of regression for sample surveys presented by Konijn [6], the notion of residuals also plays a basic role, plus either of the following assumptions:

> (i) "the population...itself constitutes a proportionate stratified sample from a (conceptually) infinitely large population of individuals with similar behavior. This makes the residual z" (equal to y - a - βx in his notation) "an ordinary random variable, in the sense of a drawing from an infinite population, uncorrelated with each x and with zero mean," and (ii) "for any given individual in the finite population the conditional mean of y given x is linear in x, and . . . the deviation z corresponding to a given individual and for a fixed value of x represents one realization among a class of potential fluctuations about this individual's conditional mean y for any fixed x",

each leading to two fifferent linear models proposed by him.

Recently Godambe and Thompson [5] introduced with much ingenuity the notion of "error" vector for their "attempt to show how the regression analysis generally used for hypothetical populations can also be validated for the actual populations commonly dealt with in statistical surveys" by starting their arguments with the assumption which they stated as follows:

> "Suppose that for every unit $i(i=1,..,N)$ of the population we have knowledge *a priori* (i.e. before any sampling is done) of the (real) value y_i associated with the unit $i(i=1,..,N)$ of some auxiliary variate y".

Their arguments for arriving at the "error" vector are different from those of Konijn of whose work they do not appear to be aware.

However, in the theory presented in section 2, these restrictive concepts of residuals or errors (restrictive, because other than the dependent random variable other random variables pertaining to the units of the sample would have to be known or to remain fixed) do not play any role. Rather, the notion of the ascribed (or hypothesized) line, surface, or hypersurface, with variables that are not random variables, passing through the centroid,* the statistics of which have desirable properties in the context of sample survey theory, plays an important role. The latter notion immediately leads to the random function, which plays a key role as illustrated in the example of section 3. Further the present theory has a different empirical basis from that of least squares. In sequel it will appear that the logical basis of this theory, characterized by an economy of principles and assumptions, is different from those of Konijn, Godambe, and Thompson.

In regard to principles, what is common between the theory presented in section 2 and least squares is the Gauss-Laplace principle of minimum variance. Because of this circumstance, in the matter of fitting a straight line (or any plane), for the case of a single stratum or universe, despite the difference in approach, the end results agree, when the line is made to pass through the estimated centroid. This agreement is evidenced by equation (27).

Thus so far in the literature of statistical theory and methods, a theory of functional relationship which directly recognizes the circumstance that all values observed on the units of a probability sample from a finite universe are random variables does not exist. The theory presented in section 2, which is fairly general, is an attempt to fill this gap, and is an elaboration of the two concise statements [9, 10] made by the writer more than eight years ago.

One illustration of the theory given in section 3 has obvious applications, for example in studying relations between income and expenditure in consumer expenditure surveys carried out by methods of statistical sampling, and doubtless other sociological surveys similarly carried out.

2. Theory

There is a finite universe U composed of N identifiable elements $\{u_i: i=1,2,...,N\}$ which are the ultimate units of sampling. These elements may be directly identifiable as single units, or identifiable indirectly as distinct members of larger units. The ℓ (≥ 2) measurable characteristics (variate values) of unit u_i are $(x_i, y_i, z_i...)$, and for the purpose of

---

* This idea was used by Boscovich as early as 1757 in fitting a line by his method of minimizing the sum of the absolute deviations [2]. I must thank Dr. Churchill Eisenhart for the gift of his paper on the work of Roger Joseph Boscovich.

this paper we assume that they can be measured without error.

A de facto functional relationship exists between the ℓ conjoint values of the respective characteristics of each element u_i as soon as the corresponding sets of finite values are defined as ordered ℓ-tuples, and the whole (finite) collection of such ℓ-tuples

(1) $\{(y_i,x_i,z_i,\ldots): i=1,2,\ldots,N\}$

defines this functional relationship completely in a mathematical sense. The way of expressing the functional relationship at (1) may be interpreted to mean that the first coordinate y, in some sense, depends on the remaining ℓ-1 corrdinates (x,z,...); this formulation is sufficient to initiate the next statement of the problem and also to suggest the possibility of casual relationships in the context of problems in the real world.

Obviously such a function is discontinuous, but the theoretical possibility that it can be satisfactorily approximated by some continuous (real) function remains open.

Generally such a (hypothetical) function approximating the de facto function, and purporting to show the relationship between the variate values of the elements of the universe may be expressed as a multivariate polynomial (of some prescribed degree with $p(\geq \ell)$ parameters, β_o, β_1, β_{p-1}, but less than the number of observations, viz,

(2) $Y=\beta_o+f(X,Z,\ldots;\beta_1,\ldots,\beta_{p-1})$,

where the domains of the ℓ-1 arguments, X,Z,..., whatever they may be, are such that each includes the respective domains of x,z,... . The ℓ-1 domains of x,z,... are defined by the extreme values relevant to each marginal distribution. The domain of Y will be determined by $\beta_o + f$. We stress here that since the (Y, X, Z,...)'s are not in one-to-one correspondence with any probability measure, they are ordinary variables despite the fact that their domains cover the domains of (y,x,z,...) which later are found to be strictly random variables. The form of the polynomial relationship between Y and X,Z,... may be suggested by previous information and/or that of the sample drawn for its estimation. The reader may now see that any of the ℓ variables may be chosen as the dependent variable so that in general ℓ different formulations similar to (2) are possible. The user of the theory will have to decide which of the ℓ formulations is revelant in the context of his problem. It may very well be that more than one formulation is revelant. Obviously the question of relevance, in problems of specific applications of the theory, is something outside the scope of a general mathematical-statistical theory. What, in sequel, will appear important for the development of the theory is that f is linear in the ℓ-1 unknown parameters.

To estimate the parameters of the functional relationship (2) we have a sample of distinct units, s, from U which is realized with positive probability p(s). To elaborate further, s is a member of a collection S, i.e. S = {s}, and p(s), a rational positive number less than unity, is defined for all s. For the sake of generality the sample design leading to s (with all its appropriate randomization procedures for the selection of units) is not specifically defined. It may be simple random sample or it may be a multi-stage sample where units at one or more stages are drawn with varying probabilities, with or without replacement. What we have are the variate values of the distinct units of s and its corresponding probability of realization, viz.

(3) $\{(y_i,x_i,z_i,\ldots): i\varepsilon s\}$ and p(s).

We conclude from the statement at (3) and the preceding explanations of this paragraph that y, x, z,... are all random variables.

It is important to note that p(s) (unlike, for example, the multivariate normal probability for correlated variates) cannot tell us anything about the nature of the relationship between the random variables y, x, z, etc., simply because it is a numerical value. However, it is through p(s) that the values of the random variables incident to the distinct units of s are revealed, and which in turn suggest the form of the (approximate) functional relationship. Logically therefore we cannot ignore p(s), or equivalently the sample design, when we attempt to infer about (1).

Now (2), as it stands, is just a hypothetical relationship between the ℓ variables, as yet unrelated to the N discrete points

(4) $(y_i,x_i,z_i,\ldots)$, $(i=1,2,\ldots,N)$

in the same ℓ-dimensional Euclidean space, which define the de facto functional relationship (1). The statistical data given at (3) in some sense represents (4). To arrive at a statistically meaningful functional relationship a conjunction of (3) and (2) needs to be effected in some way.

Heuristically we feel that some condensation of the data at (3) is required as a first step to effect conjunction with (2). It is natural to think of the centroid of these points for the purpose of condensation.

Apart from the problem of functional relationships, it is desirable from the point of a view of the known estimation theory for sample surveys, to have the true centroid of (4) and the centroid of (3), which will be an estimate of the true centroid, as close as possible in some sense.

The true centroid is given by

(5) $(\bar{\bar{y}}, \bar{\bar{x}}, \bar{\bar{z}}, \ldots)$

where

$$\bar{\bar{y}} = \sum_{i=1}^{N} y_i/N,$$

$$\bar{\bar{x}} = \sum_{i=1}^{N} x_i/N,$$

$$\bar{\bar{z}} = \sum_{i=1}^{N} z_i/N, \text{ etc.}$$

On the basis of the data at (3) let

(7) $(\hat{y}, \hat{x}, \hat{z}, \ldots)$

be estimate of (5). The estimating formulas for $\hat{y}$, $\hat{x}$, $\hat{z}$, etc. will depend on the underlying sample design. We shall require these estimates to be admissible and consistent, and more desirably to be unbiased. Because there are no best estimates in the context of sample surveys (see [4] and [7]) this is all we can do to ensure that (7) shall be "close" to (5).

To effect conjunction let the polynomial (hypersurface) (2) pass through the point (7), which is composed of estimates (based on data (3)) which are also random variables. Then we have

(8) $\hat{Y} = \hat{y} + f(X, Z, \ldots; \beta_1, \ldots, \beta_{p-1}) - f(\hat{x}, \hat{z}, \ldots; \beta_1, \ldots, \beta_{p-1})$

as the equation of the hypersurface passing through (7). Hence $\hat{Y}$ is a function of random variables or a random function for short. It is important to note that $f(X, Z, \ldots; \beta_1 \ldots, \beta_{p-1})$ in (8) is just a real function without probabilistic character since X,Z,... are not defined as random variables.

Because $\hat{Y}$ is a random function there is no way to apply the classical method of least squares for the estimation of $\beta_1, \ldots, \beta_{p-1}$. However, one of the twin principles on which least squares is founded, the Laplace-Gauss principle of minimum variance, is applicable in the sense that it allows us to determine the form of the β's which minimize the variance of the random function (8). We proceed to apply this principle. We have

(9) $$V(\hat{Y}) = V(\hat{y}) - 2 \text{ Cov } [\hat{y}, f(\hat{x}, \hat{z}, \ldots; \beta_1, \ldots, \beta_{p-1})] + V[f(\hat{x}, \hat{z}, \ldots; \beta_1, \ldots, \beta_{p-1})].$$

In determining $V(\hat{Y})$, $f(X, Z, \ldots)$ behaves as a constant by virtue of the fact that X, Z, etc., are not defined as random variables. Equation (9) reveals the interesting result that the variance of $\hat{Y}$ is constant for all $(X, Z, \ldots)$.

We note that f is a polynomial linear in the β's and also that the β's are implicit in V[f] and Cov $[\hat{y}, f]$, and both these functions which are respectively of degree two and degree one in the β's, involve the variances of $\hat{x}$, $\hat{z}$, etc., and also the relevant covariances. The value of the β's which minimize $V(\hat{Y})$ will be given by the solution of the following p-1 equations

(10) $$\frac{\partial V[(\hat{Y})}{\partial \beta_a} = \frac{\partial V[f]}{\partial \beta_a} - 2 \frac{\partial \text{ Cov}[\hat{y}, f]}{\partial \beta_a} = 0,$$

$$(a = 1, 2, \ldots p-1).$$

It is not difficult to see that the solution of the set of p-1 simultaneous equations given at (10) minimizes $V(\hat{Y})$. From (9), since the covariance function is linear in the β's, the

Hessian $\left| \frac{\partial^2 V(\hat{Y})}{\partial \beta_a \partial \beta_b} \right| = \left| \frac{\partial^2 V(f)}{\partial \beta_a \partial \beta_b} \right|$, $(a=b=1,2,\ldots,p-1)$

will obtained with variances of $\hat{x}$, $\hat{z}$, etc., making up the diagonal elements and the corresponding covariance expressions making up the off-diagonal elements. This Hessian and all its principal minors, starting with $\frac{\partial^2 V(f)}{\partial \beta_1{}^2}$, are all positive so that $V(\hat{Y})$ is minimized.

The simultaneous equations (10) involve (unknown) universe values of the variances and the covariances. Hence the solution for each β will be a ratio of functions of variances and covariances. For the sake of argument, if these variances and covariances were known, the β's obtained by solving (10) would really minimize $V(\hat{Y})$. But this cannot be realized. What can be done, as proposed in the foregoing account, is to substitute the unbiased or consistent estimates in the formal solution for each β_a.

The expressions for the estimated β's will be in the form of ratios, the respective numerators and denominators of which involve estimates of the variances and covariances. If $\hat{\beta}_a$ is an estimate of β_a, then because it is a ratio estimate

$$E(\hat{\beta}_a) \neq \beta_a, \quad (a = 1, 2, \ldots, p-1),$$

so that it is a biased estimate. We shall recur briefly to this problem in section 3. At any rate these estimated β's, despite their bias, are in some sense near-minimum values, since the values implied by their parent expressions are minimum values.

In summary, and ascribed functional relationship is estimated in the following steps:

(i) estimate the centroid of the universe,
(ii) determine the random function, i.e., the equation of the given surface

(or line) passing through this estimated centroid,

(iii) determine the variance of this random function,

(iv) determine the parameters involved so as to minimize the variance of the random function, and

(v) estimate these parameters by substituting in their expressions unbiased or consistent estimates for the variances and covariances involved, or functions thereof.

With these estimates of the parameters and the centroid, an estimate of the functional relationship is

$$(11)\quad Y' = \hat{y} + f(X, Z, \ldots; \hat{\beta}_1, \ldots; \hat{\beta}_{p-1}) - f(\hat{x}, \hat{z}, \ldots; \hat{\beta}_1, \ldots, \hat{\beta}_{p-1})$$

The expression for the variance of Y' will be extremely complex. Because of this circumstance the estimation of the variance of Y' for given (X, Z,...) for any specified sample design and functional relationship, will be difficult. A way of circumventing this difficulty through the well known technique of independent replicated samples, initiated by the late Professor Mahalanobis in the thirties in India, is discussed in section 3.

3. Example for Stratified Random Sampling

In this section we shall consider the estimation of a linear functional relationship for the case of stratified random sampling. Examples for more ramified sample designs are available but have not been included here to save space.

For the case of linear relationship with simple random sampling a brief discussion of

(i) the bias of the estimated parameter,

(ii) the estimation of the variance of a Y-value, for given X, through the technique of independent replication, and

(iii) a type of probability statement regarding the median Y-value for a given X

will be given. This discussion is intended to suggest solutions for similar problems incident to functional relationships considered with sample designs other than simple random sampling.

The universe U is subdivided into L strata each containing N_h elements (h=1,2,...,L). From stratum h, n_h (h=1,2,...;L) elements are selected with equal probabilities and without replacement at each draw, and two measurable characteristics are observed. Let (y_{hi}, x_{hi}) denote the variate values of the characteristics for the i^{th} element ($i=1,2,\ldots,n_h$) of the h^{th} stratum. We wish to estimate a linear functional relationship

$$(12)\quad Y = \beta_o + \beta_1 X.$$

We proceed as indicated in the summary paragraph of section 2. The centroid $(\bar{\bar{y}}, \bar{\bar{x}})$ is given by

$$(13)\quad \begin{cases} \bar{\bar{y}} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi} / \sum_{h=1}^{L} N_h \\ \bar{\bar{y}} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} x_{hi} / \sum_{h=1}^{L} N_h \end{cases}$$

The unbiased estimates of $\bar{\bar{y}}$ and $\bar{\bar{x}}$ are

$$(14)\quad \begin{cases} \hat{y} = \sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} y_h \\ \hat{x} = \sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} x_n \end{cases}$$

where $W_h = N_h / \sum_{h=1}^{L} N_h$ (h=1, 2,...L). Thus the equation of the line passing through the estimated centroid $(\hat{y}, \hat{x})$ will be

$$(15)\quad \hat{Y} = \hat{y} + \beta_1 (X - \hat{x}).$$

By virtue of $(\hat{y}, \hat{x})$, (15) is a random function, with variance

$$(16)\quad V(\hat{Y}) = V(\hat{y}) + \beta_1^2 V(\hat{x}) - 2\beta_1 \operatorname{Cov}(\hat{y}, \hat{x}).$$

The value of β_1 which minimizes $V(\hat{Y})$ is given by the solution of the equation $\partial V(\hat{Y})/\partial \beta_1 = 0$, that is

$$(17)\quad \beta_1 = \operatorname{Cov}(\hat{y}, \hat{x}) / V(\hat{x}).$$

This is the minimum value, since $\frac{\partial^2 V}{\partial \beta_1^2} = 2V(\hat{x}) > 0.$

Now according to standard theory

$$(18)\quad \operatorname{Cov}(\hat{y}, \hat{x}) = \sum_1^L W_h^2 \, S_{hxy} \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$$

and

$$(19)\quad V(\hat{x}) = \sum_1^L W_h^2 \, S_{hx}^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right),$$

where

$$(20)\quad S_{hxy} = \sum_{i=1}^{N_h} (x_{hi} - \bar{\bar{x}}_h)(y_{hi} - \bar{\bar{y}}_h)/(N_h - 1)$$

and

$$(21)\quad S_{hx}^2 = \sum_{i=1}^{N_h} (x_{hi} - \bar{\bar{x}}_h)^2/(N_h - 1)$$

in which $\bar{\bar{x}}_h$ and $\bar{\bar{y}}_h$ are the usual stratum means. As n_h increases, the variances of the estimates given by (14) decrease. In this connection the reader might say "what happens to β_1 if $n_h = N_h$ for all h?" Mathematically, of course, we have a situation when β_1 as expressed in terms of values given by (18) and (19) assumes an indeterminate value $\frac{0}{0}$. But here logic, the premises for which are already embedded in the basis of the problem comes to the rescue. The reply is that the need for estimating the hypothesized relationship, along the lines stated in the paper, would vanish, simply because complete knowledge of the functional relationship is already given by the de facto relationship (1), applicable to the case of two variables. The same kind of argument holds if $n_h = N_h$ for some strata. The need for determining a functional relationship would then be restricted to the remaining strata.

An estimate of β_1, is obtained by substituting unbiased estimates of Cov $(\hat{y}, \hat{x})$ and $V(\hat{x})$ in (17) and their estimation entails unbiased estimates of S_{hxy} and S^2_{hx}. For a given h, these estimates are

$$(22)\quad s_{hxy} = \sum_{i=1}^{n_h} (x_{hi}-\bar{x}_h)(y_{hi}-\bar{y}_h)/(n_h - 1)$$

and

$$(23)\quad s^2_{hx} = \sum_{i=1}^{n_h} (x_{hi}-\bar{x})^2/(n_h - 1),$$

where $\bar{x}_h$ and $\bar{y}_h$ are the usual unbiased estimates of $\bar{\bar{x}}_h$ and $\bar{\bar{y}}_h$.

Thus an estimate of the linear relationship will be

$$(24)\quad Y' = \hat{y} + \hat{\beta}_1 (X - \hat{x}).$$

where

$$(25)\quad \hat{\beta}_1 = \sum_1^L W_h^2 \frac{s_{hxy}}{n_h}\left(1 - \frac{n_h}{N_h}\right) \Big/ \sum_1^L W_h^2 \frac{s^2_{hx}}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right).$$

The formula (25) speaks for itself. We interpret the multiplying factor W_h^2 in both numerator and denominator as expressing the importance of the strata involved.

When there is proportional allocation, i.e. $n_h/\sum_1^L n_h = W_h$ for all h, then

$$(26)\quad \hat{\beta}_1' = \frac{\sum_1^L W_h s_{hxy}}{\sum_1^L W_h s^2_{hx}}$$

a result independent of sample size. Further when $W_1 = 1$, and $W_h = 0$ for $h = 2,\ldots L$, i.e. there is only one stratum, then, dropping the stratum identification subscript, we find

$$(27)\quad \hat{\beta}_1'' = \sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y}) \Big/ \sum_{i=1}^n (x_i - \bar{x})^2,$$

which is a familiar estimate obtained in the least squares method. The identity of this result with that of classical least squares regression is certainly reassuring.

By the use of the generalized k-statistics [1] it can be shown that

$$(28)\quad \frac{E(\hat{\beta}_1'') - \beta}{\beta} = \left(\frac{K_{40}}{K_{20}^2} - \frac{K_{31}}{K_{20}K_{11}}\right) \cdot \frac{(N-n)(Nn-n-N-1)}{n(n-1)N(N+1)} + 0(n^{-2}),$$

where $\beta = \sum_1^N (x_i - \bar{\bar{x}})(y_i - \bar{\bar{y}}) \Big/ \sum_1^N (x_i - \bar{\bar{x}})^2$ may be defined as the true regression coefficient and K_{40}, K_{31}, K_{20}, and K_{11} are parent bivariate k-statistics for a finite universe. Thus the bias of $\hat{\beta}_1''$ relative to the true β is of order 1/n. The proof is omitted to save space. Similar results hold for $\hat{\beta}_1$ and $\hat{\beta}_1'$ but are complicated by the weights, and additionally the sampling fractions involved in the case of $\hat{\beta}_1$.

The variance of Y', for any given X, is

$$(29)\quad V(Y') = V(\hat{y}) + X^2 V(\hat{\beta}_1) + V(\hat{\beta}_1 \hat{x}) - 2X \,\mathrm{Cov}(\hat{\beta}_1, \hat{x}) - 2\,\mathrm{Cov}(\hat{\beta}_1 \hat{x}, \hat{y}) + 2X\,\mathrm{Cov}(\hat{y}, \hat{\beta}_1),$$

Other than the expression for $V(\hat{y})$, which is similar to (19), the expressions for the remaining terms will be extremely complex and will not be exact when determined by series expansion methods. Thus, estimation of V(Y') for a given X poses a problem.

The difficulty of variance estimation can be circumvented by the technique of independent replication. Suppose the universe was not stratified and suppose we draw k independent replicated simple random samples (see [11] and [8]), each of the same size, and estimate k statistically independent lines

$$(30)\quad Y_a' = \hat{y}_a + \hat{\beta}_{1a} (X - \hat{x}_a) \quad (a = 1,2,\ldots,k)$$

according to the foregoing theory. The additional subscript in (30) identifies the k different lines. Then an estimate of the Y-ordinate, for a given X, is simply

(31) $\bar{Y}' = \frac{1}{k} \sum_{a=1}^{k} Y'_a$

with variance

(32) $V(\bar{Y}') = V(\bar{Y}')/k$

an unbiased estimate of which is

(33) $v(\bar{Y}') = \sum_{a=1}^{k} (Y'_a - \bar{Y}')^2 / \{k(k-1)\}$,

despite the complexity of the expression for $V(Y')$.

The method of independent replication also leads to a method for making probability statements about the median. If the sample size for each independent replicate is n, then we can have $\binom{N}{n}$ possible estimated lines. For a given X, the Y-ordinate will cut these lines in $\binom{N}{n}$ points, assuming that the slope of every estimated line is at angle other than 90^o. If $\binom{N}{n}$ is an even number we may define the median Y as the point midway between the $\{\binom{N}{n}/2\}^{th}$ Y-value and the $\{\binom{N}{n}/2 + 1\}^{th}$ Y-value. Recalling the results of classical nonparametric theory or arguing from first principles, we find with independent replicates, for a given X, that

(34) $P\{Y'_{least} < Y_{median} < Y'_{largest}\} = 1 - (\frac{1}{2})^{k-1}$.

Order statistics other than the extremes may also be used in making probability statements.

The reasons for restricting a statement of probability to the median of the Y's are as follows:

(A) The de facto functional relationship is restricted only to the N discrete points so that the given X (corresponding to which inference about Y is desired) may not relate to any of the true X's.

(B) Nonetheless on the supposition that X may be a possible value we wish to know the probability limits of the corresponding Y-value. For given X, the only probability distribution of the Y-values is those of the corresponding Y-ordinates of the $\binom{N}{n}$ lines generated by the randomization procedure for selecting the $\binom{N}{n}$ possible samples. The statment of probability, of necessity, must therefore be restricted to this set of Y-values (for given X), and the only exact statement of probability we can make is about the median Y-value of this distribution.

REFERENCES

[1] David, F. N., Kendall, M. G. and Barton, D. E. Symmetric Function and Allied Tables, Cambridge: Cambridge University Press 1966.

[2] Eisenhart, C., "Roger Joseph Boscovich and the Combination of Observations," Actes des Symposium International R. J. Boscovich 1961(1962), 19-25.

[3] ______________"The Meaning of "Least" in Least Squares", Journal of the Washington Academy of Sciences, 54 (1964), 24-32.

[4] Godambe, V. P., "A Unified Theory of Sampling Finite Populations", Journal of the Royal Statistical Society B, 17 (1955), 268-78.

[5] Godambe, V. P. and Thompson, M. E., "Bayes, Fiducical and Frequency Aspects of Statistical Inference in Regression Analysis in Survey-sampling", Journal of the Royal Statistical Society B, 33 (1971), 361-90.

[6] Konijn, H. S., "Regression Anaylsis in Sample Surveys", Journal of the American Statistical Association,57 (1962), 590-606.

[7] Koop, J. C., "Contributions to the General Theory of Sampling Finite Populations Without Replacement and with Unequal Probabilities", Ph.D. Thesis, North Carolina State University Library (1957). (Also in N. C. Institute of Statistics. Mimeo. Series 296, 1961).

[8] ___________On Theoretical Questions Underlying the Technique of Replicated or Interpenetrating Samples. Procedings of the Social Statistics Section, American Statistical Association (1960), 196-205.

[9] _________ Bulletin of the International Statistical Institute, 40 (1963), 1062-64, (Contributions to the discussion on T. Cunia's paper).

[10] ___________ "On the Estimation of Functional Relationships for a Finite Universe". (Abstract) Annals of Mathematical Statistics 35 (1964), 930.

[11] Lahiri, D. B. "National Sample Survey No. 5. Technical Paper on Some Aspects of the Sample Design", Sankhyā B, 25 (1954) 268-316.

[12] Sheynin, O. B., "Origin of the Theory of Errors", Nature, 211 (1966), 1003-4.