ESTIMATION OF ACCIDENTAL GAPS IN THE DATA ON SAMPLE SURVEY OF MARINE FISH LANDINGS

A. K. KESAVAN NAIR

Central Marine Fisheries Research Institute, Cochin-11

In sample surveys carried out to estimate the monthly marine fish landings, instances often arise when the relevant data in parts are lacking. A statistical method for the estimation of such gaps has been developed which seems to meet the requirement within limits.

In dealing with the data depicting time series, gaps with missing values are of common occurrence. How best these missing values could be estimated is the purpose of this communication. The method developed here, though of a general nature, was in connection with the monthly marine fish landings for a specified geographical region of India.

Continuous sample surveys are being carried out by the Central Marine Fisheries Research Institute in order to estimate, among other things, the month-wise and state-wise marine fish landings in India. To facilitate estimation, each maritime state has been divided into several zones and the collection of fish landing data is undertaken by a survey staff by making periodic visits to the landing centres. From these data, estimates are arrived at primarily for a zone. Though alternate arrangements are made fairly quickly to substitute the data collector when the regular person is absent, due to various reasons, very often it is not possible to send a substitute promptly. This results in gaps in the data.

An examination of seasonal estimates of fish landings in various zones, over a period of several years, reveals more or less similar trends. This makes it easier to estimate the missing value by comparing the monthly landings of any two years. The method given below in detail is valid only when there is only one missing value.

Method—Let T_{11}, T_{12}, T_{13} and T_{14} and T_{21}, T_{22}, T_{23}, and T_{24} be the quantities landed during the four quarters in any two years. When the landings are arranged according to quarters, for the sake of convenience, we have the following lay out:

	First *quarter*	Second *quarter*	Third *quarter*	Fourth *quarter*	Total
Year (1)	T_{11}	T_{12}	T_{13}	T_{14}	$T_{1.}$
Year (2)	T_{21}	T_{22}	T_{23}	T_{24}	$T_{2.}$
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$	$T_{.4}$	$T_{..}$

Since there is reason to believe that the classifications of the above two years do not affect the classifications due to quarters, it is possible to work out the expected values for any T_{ij} ($i=1,2$; $j=1,2,3,4$) in the very same way as in a contingency table.

The expected value of T_{ij} will then be

$$\frac{T_{i.}\ T_{.j}}{T_{..}}$$

Returning to the main problem, let it be assumed that the value T_{ij} is missing.

The totals, $T_{i.}$, $T_{.j}$ and $T_{..}$ all contain the unknown T_{ij}. Actually,

$$T_{i.} = \sum_{p\neq j=1}^{4} T_{ip} + T_{ij} \qquad \ldots\ldots\ldots (1)$$

$$T_{.j} = \sum_{q\neq i=1}^{2} T_{qj} + T_{ij} \qquad \ldots\ldots\ldots (2)$$

$$\text{and } T_{..} = \sum_{q}\sum_{p} T_{qp} + T_{ij} \quad (q\neq i,\ p\neq j) \qquad \ldots\ldots\ldots (3)$$

The first terms on the right hand side of the equations (1), (2) and (3) are known. For the sake of convenience, let these be denoted by C_1, C_2 and C_3 respectively.

The expected value of T_{ij} will then be

$$E(T_{ij}) = \frac{(C_1 + T_{ij})(C_2 + T_{ij})}{(C_3 + T_{ij})} \qquad \ldots\ldots\ldots (4)$$

Taking $E(T_{ij})$ equal to T_{ij},

$$= \frac{(C_1 + T_{ij})(C_2 + T_{ij})}{(C_3 + T_{ij})} \qquad \ldots\ldots\ldots (5)$$

Equation (5) yields a solution

$$T_{ij} = \frac{C_1 C_2}{C_3 - C_2 - C_1}$$

Illustration — Quarterly figures in one zone of Gujarat during the years 1968 and 1969 are givn below:

Quantity landed in tonnes

Year	First *quarter*	Second *quarter*	Third *quarter*	Fourth *quarter*	Total
1968	655	915	408	1092	3070
1969	567	754	356	963	2640

Let it be assumed that the value for the fourth quarter of 1969 is missing. The constants C_1, C_2 and C_3 are, then:

$C_1 = 1677$, $C_2 = 1092$ and $C_3 = 4747$,

and hence the estimate for the quarter turns out to be 926.

The calculated value comes very close to the actual value of 963 tonnes. The error computed as a percentage of the actual value is about 4%.

In a similar manner the estimates for the first, second and third quarters can also be calculated as 562, 801 and 350 tonnes respectively, all of which are very close to the actual values. The error as a percentage of the actual value is the highest for the estimate of the second quarter, but it does not exceed 6%.

By taking $E(T_{ij}) = T_{ij}$, the most probable value of T_{ij} is not searched. This introduces a certain amount of bias in the estimate. But so long as the values contributed to the chi-square by the available three quarters remain small, the bias involved does not seem to be serious. The method has the advantage of easy computation. Perhaps the distribution of chi-square in contingency may be of help to obtain a statistical estimate of the error. An unbiased estimate is however obtainable by the modified chi-square minimum method, which, with respect to the problem at hand, is identical with the "maximum likelihood method" (Cramer, 1951). But the modified chi-square minimum method leads to the solution of a cumbersome cubic function. Though this can be solved, the computation involved in dealing with large numerical values is rather laborious.

I am grateful to Dr. S. Z. Qasim and Shri. S. K. Banerji for critically going through the manuscript and for suggesting improvements.

CRAMER, H. 1951. *Mathematical methods of statistics*, 575 pp. Princeton University Press, Princeton.