PRODUCT ESTIMATORS IN SAMPLE SURVEYS

A Dissertation

by

JAMES ERVIN GINN

Submitted to the Graduate College of
Texas A&M University in
partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

May 1973

Major Subject: Statistics

PRODUCT ESTIMATORS IN SAMPLE SURVEYS

A Dissertation

by

JAMES ERVIN GINN

Approved as to style and content by:

H. O. Hartley
(Chairman of Committee)

H. O. Hartley
(Head of Department)

Charles E. Gates
(Member)

W. B. Smith
(Member)

(Member)

(Member)

May 1973

ABSTRACT

Product Estimators in Sample Surveys. (May 1973)

James Ervin Ginn, B.A., Clark College;

M.S., Atlanta University

Directed by: Dr. H. O. Hartley

In the theory of sampling from finite populations, the well known concept of "ratio estimator" is an extremely useful and widely used device.

This dissertation is concerned with developing an analogous concept in the form of a "product estimator" constituting as it were the product of two or more random variables X•Y•Z•••.

More specifically this dissertation covers the development of formulas for expectations and variances of products of two and three factors in both, samples from finite and infinite populations. The so-called k-statistics, computed from simple random samples, are then employed for the estimation of the components of the variance formulas. The concept of "an unbiased product estimator" is introduced and exact variance formulas derived. An application of the concepts developed to a problem in crop estimation is illustrated by data supplied by the Statistical Reporting Service (U.S.D.A.).

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. Herman O. Hartley, my Committee Chairman, who introduced me to product estimation and guided me through this research. I would like to thank each and every professor in the Institute of Statistics, for each of you have helped me one time or another. To Claude McHenry, thank you for the writing of the computer program which was used to carry out the computations. I would also like to thank Dr. A. M. Kshirsagar for his assistance in the research department and Dr. Hurshell H. Hunt for the help he gave me and listening to my problems. I, also, wish to express my sincere appreciation to Dr. William B. Smith, who guided me through my course of study at TAMU, for all the help and assistance he gave me during my stay at TAMU. My gratitude is extended to all other members for their willingness to serve on my committee: Dr. A. M. Mayyasi, Dr. C. E. Gates, Mr. B. C. Moore and Dr. A. R. Burgess. Mrs. Tina Fields was patient and efficient in typing this dissertation within a short period of time. My thanks go to her.

I am grateful to my wife, Margaret, and my son, James, for their patience, understanding, and sacrifices in order that this undertaking could be completed. To my mother, Mrs. Idell Patterson, I would like to say, thank you, as I have said many times before, for all the sacrifices you made, during the years when it was almost impossible (money-wise) to send me to college.

I am further grateful to Texas A&M University for partial support of my education at the University and to Texas Southern University for their encouragement and earlier financial support.

TABLE OF CONTENTS

SECTION PAGE

1. INTRODUCTION

In the theory of sampling from finite populations, the well known concept of "ratio estimator" is an extremely useful and widely used device (see e.g., Cochran, W. G. [1]). This dissertation is concerned with developing an analogous concept in the form of a "product estimator" constituting as it were the product of two or more random variables $X \cdot Y \cdot Z \cdots$. In analogy to the use of ratio estimation two main purposes for the use of a product estimator may be distinguished:

(A) The objective of estimating the expectation of a product of two variables $X_i \cdot Y_i$ in a population of N units (i = 1, 2, ..., N). For example in agricultural surveys we may have collected information on

X_i = acreage of a crop within a segment i

Y_i = yield per acres of that crop in segment i on a subsample basis.

Here the parameter of interest is an estimate of the total yield of the crop $P = \sum_i X_i Y_i$ for the population of segments and not the product $\overline{X}\,\overline{Y}$.

Moreover numerous examples arise in which separate survey information is available on the two factors, X_i and Y_i, and the theory of the product estimator to be developed will provide the

---

The citations on the following pages follow the style of the *Journal of the American Statistical Association*.

means of utilizing such separate data information.

(B) The estimation of the mean value of one of the factors given a true value of the mean value of the other factor. For the analogous situation in ratio estimation, (see e.g., Goodman, L. A., and Hartley, H. O. [2]). While the analogy to this case (B) in ratio estimation is the predominant one, cases in which this situation arises in conjunction with the product estimator are less frequent. It is for this purpose that "unbiased product estimators of $\overline{Y}$" are developed in section 6. However, the fundamental formulas developed in sections 3 to 5 aid both objectives (A) and (B).

2. DEFINITION OF PRODUCT ESTIMATORS AND THEIR ANALOGY TO RATIO ESTIMATORS

2.1 Ratio Estimators

Ratio estimators are frequently employed in sample surveys when estimating the population mean $\bar{Y}$ of a variable y with the help of the known population mean $\bar{X}$ of a variable x that is positively correlated with y. In the case of a simple random sample of n pairs (x_i, y_i) drawn from a population of N pairs, various ratio estimators of $\bar{Y}$ can be formed. The following two estimators, which are the usual ratio estimators, are the "ratio of means estimator"

$$\tilde{y} = \bar{X} \left(\frac{\bar{y}}{\bar{x}}\right) \qquad (2.1.1)$$

and the "mean of ratios estimator"

$$\hat{y} = \bar{X}\,\bar{r} \qquad (2.1.2)$$

where $\bar{y}$, $\bar{x}$ and $\bar{r}$ are the arithmetic means of the sample of y_i, x_i and r_i respectively, where $r_i = \frac{y_i}{x_i}$ (see e.g., W. G. Cochran [1]).

Estimators (2.1.1) and (2.1.2) are useful in situations where the x's and y's are positively correlated.

2.2 The Uses of Product Estimators in Random Samples

In analogy to ratio estimators we distinguish two cases:

(a) Given a random sample of n pairs (x_i, y_i) from a population of N pairs it is required to estimate the population mean of the products Ex_iy_i, and

(b) From the sample in (a) it is required to estimate the mean value of y_i, i.e., $Ey_i = \bar{Y}$ given $Ex_i = \bar{X}$.

Task (a) could of course employ standard finite sample theory applied to the product

$$p_i = x_iy_i \tag{2.2.1}$$

where $P_i = x_iy_i$ is regarded as the characteristic attached to unit i. However, the use of x_i and y_i in conjunction with P_i may be required because of the following three reasons:

(1) It may be required to estimate $\bar{X} \cdot \bar{Y}$ rather than Ep_i

(2) It may be required to estimate $\bar{Y}$ with $\bar{X}$ known and the coefficient of variation of P_i is less than the coefficient of variation of y_i.

Case (b) requires the study of "product estimators" of the following form, which are defined in analogy to the ratio estimators (1) and (2):

The mean of products estimator

$$\tilde{y} = \frac{1}{n} \sum_i x_i y_i / \bar{X} = \bar{p}/\bar{X} \qquad (2.2.2)$$

and the product of means estimator

$$\hat{y} = \bar{x}\,\bar{y}/\bar{X} \qquad (2.2.3)$$

Clearly the sample theory for $\tilde{y}$ can be based on that for the products p_i. However since it is of importance to compare the performance of $\tilde{y}$ with that of the sample mean $\bar{y}$ it will be necessary to express the expectation and variance of $\tilde{y}$ in terms of the moments and cross product moments of the population of x_i and y_i. This is done in section 3. In the case of the product of means estimator $\hat{Y}$ it is also necessary to use such a reduction to the moments and product moments of the population of x_i, y_i particularly as only one single product $\bar{x}\,\bar{y}$ is available for estimation.

(3) Associated values of x_i and y_i are only available for subsample and additional information is available for y_i and/or x_i separately.

This is dealt with in section 7, and an example of this is discussed in section 8.

3. EXPECTATION AND VARIANCE OF THE PRODUCT $x_i y_i$ IN TERMS OF PRODUCT MOMENTS OF x_i AND y_i

3.1 The General Case (X_i and Y_i Are Not Independent)

In this section we start with the case of two random variables and will extend our results to three random variables in section 4. Let x_i and y_i be two random variables from a population of size N not necessarily independent. Then we can write

$$E(x_i y_i) = E(x_i)\ E(y_i) + \mathrm{Cov}(x_i,\ y_i) \tag{3.1.1}$$

Note: We employ the symbols $x_i y_i$ to denote random variables in a finite population since (following Cochran) the symbols XY denote the population totals.

Where

$$\mathrm{Cov}(x_i,\ y_i) = N^{-1} \sum_{i=1}^{N} (x_i - \overline{X})(y_i - \overline{Y}) \tag{3.1.2}$$

Let

$$\varepsilon x_i = x_i - \overline{X} \quad \text{and} \quad \varepsilon y_i = y_i - \overline{Y} \tag{3.1.3}$$

which implies

$$x_i y_i = (\varepsilon x_i + \overline{X})(\varepsilon y_i + \overline{Y}) \tag{3.1.4}$$

where

$$\mathrm{Var}(x_i y_i) = E[x_i y_i - E(x_i y_i)]^2$$

$$= E[\varepsilon x_i y_i + \varepsilon x_i \overline{Y} + \overline{X}\varepsilon y_i - \mathrm{Cov}(x_i, y_i)]^2$$

After expanding and collecting terms we can write

$$\mathrm{Var}(x_i y_i) = \overline{X}^2 \mathrm{Var}(y_i) + \overline{Y}^2 \mathrm{Var}(x_i) + E(\varepsilon x_i^2 \varepsilon y_i^2)$$

$$- \mathrm{Cov}^2(x_i, y_i) + 2 \overline{Y} E(\varepsilon x_i^2 \varepsilon y_i)$$

$$+ 2 \overline{X} E(\varepsilon x_i \varepsilon y_i^2) + 2 \overline{X} \overline{Y} \mathrm{Cov}(x_i, y_i) \qquad (3.1.5)$$

Where

$$E(\varepsilon x_i^2 \varepsilon y_i^2) = N^{-1} \sum_{i=1}^{N} (x_i - \overline{X})^2 (y_i - \overline{Y})^2 \qquad (3.1.6)$$

and corresponding formulas hold for the expectations. Equation (3.1.5) is the exact formula, when x_i and y_i are not necessarily independent. This agrees with the results found by Haldane [3] , and Goodman [4].

3.2 The Case Where x_i and y_i are Independent, and the Approximation Formula

Let x_i and y_i be two independent random variables. Then we may write equation (3.1.5) as

$$Var(x_i y_i) = \overline{X}^2 Var(y) + \overline{Y}^2 Var(x) + Var(x) Var(y) \qquad (3.2.1)$$

Introducing the coefficient of variation

$$C.V.(x_i) = \frac{\sigma_x}{\overline{X}} \qquad (3.2.2)$$

We can write

$$Var(x_i) = \overline{X}^2 (C.V.(x_i))^2 \qquad (3.2.3)$$

and

$$Var(y_i) = \overline{Y}^2 (C.V.(y_i))^2 \qquad (3.2.4)$$

Now (3.2.1) can be written

$$Var(x_i y_i) = \overline{X}^2 \overline{Y}^2 (C.V.(y_i))^2 + \overline{Y}^2 \overline{X}^2 (C.V.(x_i))^2$$

$$+ \overline{X}^2 (C.V.(x_i))^2 \overline{Y}^2 (C.V.(y_i))^2 \qquad (3.2.5)$$

$$= \overline{X}^2 \overline{Y}^2 \, [C.V.(y_i)^2 + C.V.(x_i)^2 + C.V.(x_i)^2 C.V.(y_i)^2]$$

If $C.V.(x_i)$ or $C.V.(y_i)$ is small, then (3.2.5) can be approximated by

$$\tilde{\mathrm{Var}}(x_i y_i) = \bar{X}^2\bar{Y}^2[\mathrm{C.V.}(x_i)^2 + \mathrm{C.V.}(y_i)^2] \tag{3.2.6}$$

How good is the approximation? To answer this question we consider

$$\frac{\mathrm{Var}(x_i y_i) - \tilde{\mathrm{Var}}(x_i y_i)}{\mathrm{Var}(x_i y_i)} \tag{3.2.7}$$

the relative inaccuracy of the approximation where

$$\frac{\mathrm{Var}(x_i y_i) - \tilde{\mathrm{Var}}(x_i y_i)}{\mathrm{Var}(x_i y_i)} = \frac{\mathrm{C.V.}(x_i)^2\ \mathrm{C.V.}(y_i)^2}{\mathrm{C.V.}(x_i)^2+\mathrm{C.V.}(y_i)^2+\mathrm{C.V.}(x_i)^2\mathrm{C.V.}(y_i)^2}$$

$$= \frac{1}{\mathrm{C.V.}(x_i)^{-2}+\mathrm{C.V.}(y_i)^{-2}+1} \tag{3.2.8}$$

From (3.2.8) we see that, if both $\mathrm{C.V.}(x_i)$ and $\mathrm{C.V.}(y_i)$ are small, then $\mathrm{C.V.}(x_i)^{-2} + \mathrm{C.V.}(y_i)^{-2}$ will be relatively large, and the relative inaccuracy will be small.

Therefore the approximation is good if both $\mathrm{C.V.}(x_i)$ and $\mathrm{C.V.}(y_i)$ are small.

3.3 Expectation and Variance of Mean of Products Estimator $\tilde{y}$

These parameters are an immediate consequence of the fact that from (2.2.2) we have $\tilde{y} = \overline{p}/\overline{X}$ and accordingly from (3.1.1)

$$E(\tilde{y}) = \bar{X}^{-1} E(\bar{p}) = \bar{X}^{-1} E x_i y_i$$

$$= \bar{X}^{-1} \{\bar{X}\,\bar{Y} + \mathrm{Cov}(x_i, y_i)\} \qquad (3.3.1)$$

and from (3.1.5)

$$\mathrm{Var}\ \tilde{y} = n^{-1} \bar{X}^{-2}(1 - \frac{n}{N}) \frac{N}{N-1} \mathrm{Var}(x_i y_i) \qquad (3.3.2)$$

where $\mathrm{Var}(x_i y_i)$ is given by (3.1.5).

3.4 Expectation and Variance of the Product of Means Estimator

Using again the general formulas (3.1.1) and (3.1.5) we apply these to the finite population of $\binom{N}{n}$ pairs of means $\bar{x}$, $\bar{y}$. Accordingly

$$E(\hat{y}) = E\ \{\bar{X}^{-1}\ \bar{x}\ \bar{y}\}$$

$$= \bar{X}^{-1}\ \ \bar{X}\ \bar{Y} + \mathrm{Cov}(\bar{x}, \bar{y}) \qquad (3.3.3)$$

$$= \bar{Y} + \bar{X}^{-1} \frac{1}{n}(1 - \frac{n}{N}) \frac{N}{N-1} \mathrm{Cov}(x_i, y_i)$$

In order to obtain a formula for $\mathrm{Var}(\hat{y})$ we may confine ourselves to the three leading terms in (3.1.5), namely, $\bar{X}^2\ \mathrm{Var}(\bar{y})$, $\bar{Y}^2\ \mathrm{Var}(\bar{x})$

and $2\bar{X}\,\bar{Y}\,\mathrm{Cov}(\bar{x}, \bar{y})$. Accordingly we obtain the approximate formula

$$\mathrm{Var}(\hat{y}) \doteq \bar{X}^{-2} \frac{1}{n} (1 - \frac{n}{N}) \frac{N}{N-1} \{\bar{X}^2 \mathrm{Var}(y_i) + \bar{Y}^2 \mathrm{Var}(x_i)$$

$$+ 2\bar{X}\,\bar{Y}\,\mathrm{Cov}(x_i, y_i)\}$$

$$\doteq \bar{X}^{-2} \frac{1}{n} (1 - \frac{n}{N}) \frac{N}{N-1} \mathrm{Var}(\bar{X}\, y_i + \bar{Y}\, x_i) \qquad (3.3.4)$$

The justification of confining (3.3.4) to the leading terms arises from the fact that normally C.V.$(\bar{x})$ and C.V.$(\bar{y})$ are small and the three terms retained are those of second order in the C.V.'s.

4. EXACT FORMULA FOR VARIANCE OF THREE RANDOM VARIABLES

Let u, v, and w be three random variables.

Define

$$\delta u_i = \frac{u_i - \overline{U}}{\overline{U}} \tag{4.1}$$

$$\delta v_i = \frac{v_i - \overline{V}}{\overline{V}} \tag{4.2}$$

and

$$\delta w_i = \frac{w_i - \overline{W}}{\overline{W}} \tag{4.3}$$

Then we can write

$$u_i = \overline{U}\ (\delta u_i + 1) \tag{4.4}$$

$$v_i = \overline{V}\ (\delta v_i + 1) \tag{4.5}$$

and

$$w_i = \overline{W}\ (\delta w_i + 1) \tag{4.6}$$

Therefore

$$u_i v_i w_i = \overline{U}\,\overline{V}\,\overline{W}\,(\delta u_i + 1)(\delta v_i + 1)(\delta w_i + 1) \tag{4.7}$$

Accordingly

$$\begin{aligned} Eu_i v_i w_i &= \overline{U}\,\overline{V}\,\overline{W}\,\{1 + \frac{Cov(u_i,\ v_i)}{\overline{U}\,\overline{V}} + \frac{Cov(u_i,\ w_i)}{\overline{U}\,\overline{W}} \\ &\quad + \frac{Cov(v_i,\ w_i)}{\overline{V}\,\overline{W}} + E\delta u_i\ \delta v_i\ \delta w_i\} \\ &= \overline{U}\,\overline{V}\,\overline{W} + \overline{W}\,Cov(u_i,\ v_i) + \overline{V}\,Cov(u_i,\ w_i) \\ &\quad + \overline{U}\,Cov(v_i,\ w_i) + E(u_i-\overline{U})(v_i-\overline{V})(w_i-\overline{W}) \end{aligned} \tag{4.8}$$

Turning now to the variance we obtain

$$\begin{aligned} Var(u_i v_i w_i) &= E\ \{\overline{U}\,\overline{V}\,\overline{W}\,(1+\delta u_i)(1+\delta v_i)(1+\delta w_i) \\ &\quad - Eu_i v_i w_i\}^2 \end{aligned} \tag{4.9}$$

where $Eu_i v_i w_i$ is given by (4.8). After considerable algebra we obtain

$$\mathrm{Var}(u_i v_i w_i) = \overline{U}^2\ \overline{V}^2\ \mathrm{Var}(w_i) + 2\overline{U}^2\ \overline{V}\ E(\varepsilon_{v_i}\ \varepsilon^2_{w_i}) + \overline{U}^2\ \overline{W}^2\ \mathrm{Var}(v_i)$$

$$+ 2\overline{U}^2\ \overline{W}\ E(\varepsilon_{w_i}\ \varepsilon^2_{v_i}) + \overline{U}^2\ E(\varepsilon^2_{v_i}\ \varepsilon^2_{w_i})$$

$$+ 4\overline{U}\ \overline{V}\ \overline{W}\ E(\varepsilon_{u_i}\ \varepsilon_{v_i}\ \varepsilon_{w_i}) + 2\overline{U}\ \overline{V}\ E(\varepsilon_{u_i}\ \varepsilon_{v_i}\ \varepsilon^2_{w_i})$$

$$+ \overline{U}\ \overline{V}\ \overline{W}\ E(\varepsilon_{u_i}\ \varepsilon_{v_i}\ \varepsilon_{w_i})^2 - 2\overline{U}\ \overline{W}\ \mathrm{Cov}(u_i, v_i)\ E(\varepsilon_{v_i}\ \varepsilon_{w_i})$$

$$+ 2\overline{U}\ \overline{W}^2\ E(\varepsilon_{u_i}\ \varepsilon^2_{v_i}) + 4\overline{U}\ \overline{W}\ E(\varepsilon_{u_i}\ \varepsilon^2_{v_i}\ \varepsilon_{w_i})$$

$$+ 2\overline{U}\ E(\varepsilon_{u_i}\ \varepsilon^2_{v_i}\ \varepsilon^2_{w_i}) + \overline{V}^2\ \overline{W}^2\ \mathrm{Var}(u_i)$$

$$+ 2\overline{V}^2\ \overline{W}\ E(\varepsilon^2_{u_i}\ \varepsilon_{w_i}) + 2\overline{V}\ \overline{W}^2\ E(\varepsilon^2_{u_i}\ \varepsilon_{v_i})$$

$$+ 2\overline{V}\ \overline{W}\ E(\varepsilon^2_{u_i}\ \varepsilon_{v_i}\ \varepsilon_{w_i}) + 2\overline{V}\ E(\varepsilon^2_{u_i}\ \varepsilon_{v_i}\ \varepsilon^2_{w_i})$$

$$- 2\overline{U}\ \mathrm{Cov}(v_i, w_i)\ E(\varepsilon_{u_i}\ \varepsilon_{v_i}\ \varepsilon_{w_i}) + 2\overline{W}^2\ \mathrm{Cov}(u_i, v_i)$$

$$+ 2\overline{W}\ E(\varepsilon^2_{u_i}\ \varepsilon_{w_i}\ \varepsilon^2_{v_i}) + E(\varepsilon^2_{u_i}\ \varepsilon^2_{v_i}\ \varepsilon^2_{w_i}) - 2\overline{U}^2 \mathrm{Cov}^2(v_i, w_i)$$

$$- 2\overline{U}\ E(\varepsilon_{u_i}\ \varepsilon_{v_i}\ \varepsilon_{w_i})\ \mathrm{Cov}(v_i, w_i)$$

$$- 2\overline{V}\ \overline{W}\ \mathrm{Cov}(u_i, v_i) \mathrm{Cov}(u_i, w_i) - 2\overline{U}\ \overline{V}\ \overline{W}^2\ \mathrm{Cov}^2(u_i, v_i)$$

$$- 2\overline{V}^2 \, Cov(u_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{w_i}) - 2\overline{U}\,\overline{V} \, Cov(v_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{w_i})$$

$$- \overline{V} \, E(\varepsilon_{u_i} \varepsilon_{w_i}) \, E(\varepsilon_{u_i} \varepsilon_{v_i} \varepsilon_{w_i}) - \overline{V}\,\overline{U} \, Cov(u_i,w_i) \, E(\varepsilon_{v_i} \varepsilon_{w_i})$$

$$- 2\overline{W}^2 \, Cov(u_i,v_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i}) - \overline{V}\,\overline{W} \, Cov(u_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i})$$

$$- 2\overline{U}\,\overline{W} \, Cov(v_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i}) + 2\overline{U}\,\overline{V}^2 \, E(\varepsilon_{u_i} \varepsilon_{w_i}^2)$$

$$+ \overline{W}^2 \, E(\varepsilon_{u_i}^2 \varepsilon_{v_i}^2) + 2\overline{U}\,\overline{V}\,\overline{W} \, E(\varepsilon_{u_i} \varepsilon_{v_i} \varepsilon_{w_i})$$

$$+ 2\overline{U}\,\overline{V}^2\,\overline{W} \, E(\varepsilon_{u_i} \varepsilon_{w_i}) + 2\overline{U}^2\,\overline{V}\,\overline{W} \, E(\varepsilon_{v_i} \varepsilon_{w_i}) + \overline{V}^2 \, E(\varepsilon_{u_i}^2 \varepsilon_{w_i}^2)$$

$$+ \overline{W}^2 \, Cov(u_i,v_i) + \overline{W}\,\overline{V} \, Cov(u_i,v_i) \, Cov(u_i,w_i)$$

$$+ \overline{W}\,\overline{U} \, Cov(u_i,v_i) \, Cov(v_i,w_i) + \overline{W} \, Cov(u_i,v_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i} \varepsilon_{w_i})$$

$$+ \overline{V}\,\overline{W} \, Cov(u_i,w_i) \, Cov(u_i,v_i) + \overline{V}^2 \, Cov^2(u_i,w_i)$$

$$+ \overline{V}\,\overline{U} \, Cov(u_i,w_i) \, Cov(v_i,w_i) + \overline{V} \, Cov(u_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i} \varepsilon_{w_i})$$

$$+ \overline{U}\,\overline{W} \, Cov(v_i,w_i) \, Cov(u_i,v_i) + \overline{U}\,\overline{V} \, Cov(v_i,w_i) \, Cov(u_i,w_i)$$

$$+ \overline{U}^2 \, Cov^2(v_i,w_i) + \overline{U} \, Cov(v_i,w_i) \, E(\varepsilon_{u_i} \varepsilon_{v_i} \varepsilon_{w_i})$$

$$+ \overline{W}\, Cov(u_i,v_i)\, E(\varepsilon_{u_i}\varepsilon_{v_i}\varepsilon_{w_i}) + \overline{V}\, Cov(u_i,w_i)\, E(\varepsilon_{u_i}\varepsilon_{v_i}\varepsilon_{w_i})$$

$$+ \overline{U}\, Cov(u_i,w_i)\, E(\varepsilon_{u_i}\varepsilon_{v_i}\varepsilon_{w_i}) + [E(\varepsilon_{u_i}\varepsilon_{v_i}\varepsilon_{w_i})]^2 \qquad (4.10)$$

For the special case (frequently encountered when one of the factors (say u_i) is independent of the other two (v_i,w_i)) formula (4.10) reduces to

$$Var(u_iv_iw_i) = \overline{U}^2\,\overline{V}^2\, Var(w_i) + \overline{U}^2\,\overline{W}^2\, Var(v_i) + \overline{V}^2\,\overline{W}^2\, Var(u_i)$$

$$+ \overline{U}^2\, E(\varepsilon^2_{v_i}\varepsilon^2_{w_i}) + \overline{W}^2\, E(\varepsilon^2_{u_i}\varepsilon^2_{v_i}) + \overline{V}^2\, E(\varepsilon^2_{u_i}\varepsilon^2_{w_i})$$

$$+ 2\overline{U}^2\,\overline{V}\, E(\varepsilon_{v_i}\varepsilon^2_{w_i}) + 2\overline{U}^2\,\overline{V}\,\overline{W}\, E(\varepsilon_{v_i}\varepsilon_{w_i}) + 2\overline{W}\, Var(u_i) E(\varepsilon_{w_i}\varepsilon^2_{v_i})$$

$$- Var\, u_i\, [E(\varepsilon_{v_i}\varepsilon_{w_i})]^2 - \overline{U}^2\, Cov^2(v_i,w_i)$$

$$+ 2\overline{V}\,\overline{W}\, Var(u_i)\, E(\varepsilon_{v_i}\varepsilon_{w_i}) + 2\overline{V}\, Var(u_i)\, E(\varepsilon_{v_i}\varepsilon^2_{w_i})$$

$$+ 2\overline{U}^2\,\overline{W}\, E(\varepsilon_{w_i}\varepsilon^2_{v_i}) + Var(u_i)\, E(\varepsilon^2_{v_i}\varepsilon^2_{w_i}) \qquad (4.11)$$

where $\varepsilon_{v_i} = v_i - \overline{V}$, $\varepsilon_{u_i} = u_i - \overline{U}$, and $\varepsilon_{w_i} = w_i - \overline{W}$. Using equation (3.2.1) where u_i is independent of v_i and w_i we can write

$$\mathrm{Var}\ (u_i v_i w_i) = \overline{U}^2\ \mathrm{Var}(v_i w_i) + \overline{V}^2\ \overline{W}^2\ \mathrm{Var}(u)$$

$$+ \mathrm{Var}(u_i)\ \mathrm{Var}(v_i w_i) \qquad (4.12)$$

Upon substituting formula for $\mathrm{Var}(v_i w_i)$, which is obtained from equation (3.1.5) into equation (4.12) we see that equation (4.11) checks.

5. ESTIMATION OF VARIANCE FORMULAS

In this section, we develop formulas for the estimation of the parameter combinations occurring in the variance formulas developed in the preceding section. However, it will be appreciated that these formulas will be extremely lengthy. Therefore, we give in this section only the formulas that apply to the special case when all factors are independent. Such formulas are extremely simple. However, they will not in general be adequate if the factors are dependent upon one another.

It suffices here to give the principles on which the general unbiased estimation will be based, leaving the spelling out of the lengthy formula for the general case to the Appendix. We refer to the general variance formulas (3.1.5) and (4.10).

Referring first to the case of two factors all additive terms in formula (3.1.5) are power products of moments or power moments of the joint distribution of the x_i and y_i. Accordingly they can all be expressed in terms of power products of the finite population κ statistics. Accordingly, such power products can be unbiasedly estimated by the finite population k-statistics and their power products. Since it will be appreciated that the resulting expressions will become extremely involved so as to make their practical use almost prohibitive, we shall employ the customary approximation procedure frequently used in the theory of k-statistics. We shall use the k-statistics only to provide unbiased estimates of the following parameter functions.

$$\overline{XY},\ \overline{X}^2,\ \overline{Y}^2,\ \text{Cov } y_i x_i,\ E\ \varepsilon x_i^{\ 2}\ \varepsilon y_i^{\ 2},\ E\ \varepsilon x_i^{\ 2}\ \varepsilon y_i,\ E\ \varepsilon x_i\ \varepsilon y_i^{\ 2}$$

These unbiased estimates will then be substituted in the terms of equation (3.1.5) resulting in approximately unbiased estimates of that variance formula in the following sense. Wherever an additive term consists of a product of two or three of the terms listed above the covariances between the estimates of these terms will be ignored. A corresponding procedure will be used for the terms in formula (4.10) using the joint distribution of the three variables u_i, v_i, w_i.

5.1 Estimation of Variance Formula for Product of Two Random Variables (Independence Assumed)

In the case of two independent factors x_i and y_i formula (3.1.5) reduced to formula (3.2.1) and it is obvious that the following statistics estimate unbiassedly the parameter functions involved.

$$\begin{aligned} &E\ \{\bar{x}^2 - \frac{s_x^{\ 2}}{n}\frac{(N-n)}{N}\} = \overline{X}^2 \\ &E\ \{\bar{y}^2 - \frac{s_y^{\ 2}}{n}\frac{(N-n)}{N}\} = \overline{Y}^2 \\ &E\ s_x^{\ 2}\ \frac{N-1}{N} = \text{Var } x_i \\ &E\ s_y^{\ 2}\ \frac{N-1}{N} = \text{Var } y_i \end{aligned} \qquad (5.1.1)$$

where

$$s_x^2 = (n - 1)^{-1} \Sigma(x_i - \bar{x})^2 , \quad s_y^2 = (n - 1)^{-1} \Sigma(y_i - \bar{y})^2 . \qquad (5.1.2)$$

Accordingly, an unbiassed estimate of Var $x_i y_i$ is given by

$$\begin{aligned} \text{Var } x_i y_i = & [\bar{x}^2 - \frac{s_x^2}{n} \frac{(N-n)}{N}] \frac{N-1}{N} s_y^2 \\ & + [\bar{y}^2 - \frac{s_y^2}{n} \frac{(N-n)}{N}] \frac{N-1}{N} s_x^2 \qquad (5.1.3) \\ & + (\frac{N-1}{N})^2 s_x^2 s_y^2 \end{aligned}$$

5.2 Estimation of Variance Formula for Product of Three Random Variables (Independence Assumed)

In the case of three independent factors u_i, v_i, w_i formula (4.10) reduces to

$$\begin{aligned} \text{Var } (u_i\, v_i\, w_i) = & \bar{U}^2\bar{V}^2 \text{ Var } w_i + \bar{U}^2\bar{W}^2 \text{ Var } v_i + \bar{V}^2\bar{W}^2 \text{ Var } u_i \\ & + \bar{U}^2 \text{ Var } v_i \text{ Var } w_i + \bar{V}^2 \text{ Var } u_i \text{ Var } w_i + \qquad (5.2.1) \\ & + \bar{W}^2 \text{ Var } u_i \text{ Var } v_i + \text{Var } u_i \text{ Var } v_i \text{ Var } w_i \;. \end{aligned}$$

Using the same principles as in section (5.1) we obtain an unbiassed estimate of Var$(u_i v_i w_i)$ by making the following substitutions in (5.2.1)

For	Substitute
$\overline{U}^2$	$\bar{u}^2 - \frac{s_u^2}{n} \frac{(N-n)}{N}$
$\overline{V}^2$	$\bar{v}^2 - \frac{s_v^2}{n} \frac{(N-n)}{N}$
$\overline{W}^2$	$\bar{w}^2 - \frac{s_w^2}{n} \frac{(N-n)}{N}$
Var u_i	$s_u^2 \ (N - 1)/N$
Var v_i	$s_v^2 \ (N - 1)/N$
Var w_i	$s_w^2 \ (N - 1)/N$

(5.2.2)

5.3 Estimation of Variance of $\bar{x} \cdot \bar{y}$ in Case of Independence

Using formula (3.2.1) for $x_i = \bar{x}$ based on n(x) observations and $y_i = \bar{y}$ based on n(y) observations, we obtain

$$\text{Var } (\bar{x} \cdot \bar{y}) = \overline{X}^2 \text{ Var } \bar{y} + \overline{Y}^2 \text{ Var } \bar{x} + \text{Var } \bar{x} \text{ Var } \bar{y} \qquad (5.3.1)$$

Therefore an unbiassed estimate of Var $\bar{x} \cdot \bar{y}$ when n(x) = n(y) = n, is given by,

$$\widehat{\text{Var}}(\bar{x}\cdot\bar{y}) = [\bar{x}^2 - \frac{s_x^{\ 2}}{n}\frac{(N-n)}{N}]\frac{s_y^{\ 2}}{n}(1 - \frac{n}{N})$$

$$+ [\bar{y}^2 - \frac{s_y^{\ 2}}{n}\frac{(N-n)}{N}]\frac{s_x^{\ 2}}{n}(1 - \frac{n}{N}) \qquad (5.3.2)$$

$$+ s_x^{\ 2}\, s_y^{\ 2}\, \frac{(N-n)^2}{n^2 N^2}$$

which clearly reduces to (5.1.3) when n = 1.

The variance estimation for a triple product is analogous.

5.4 Estimate of Variance of the Mean of a Product $\overline{x_i y_i}$

We obviously have that

$$\text{Var}\ \overline{x_i y_i} = \frac{1}{n}(1 - \frac{n}{N})\frac{N}{N-1}\ \text{Var}\ x_i y_i \qquad (5.4.1)$$

Accordingly, we obtain an unbiassed estimate of Var $\overline{x_i y_i}$ from (5.1.3) as

$$\widehat{\text{Var}}\ \overline{x_i y_i} = \{\widehat{\text{Var}}\ x_i y_i\}\ \frac{1}{n}(1 - \frac{n}{N})\ \frac{N}{N-1} \qquad (5.4.2)$$

5.5 Estimation of the Variance of the Product of Means Estimator (Case of Dependance)

To estimate the variance of the product of means estimator we may confine ourselves to the three leading terms in (3.1.5) namely $\bar{X}^2$ Var($\bar{y}$), $\bar{Y}^2$ Var($\bar{x}$), and 2 $\bar{X}$ $\bar{Y}$ Cov($\bar{x}$, $\bar{y}$). Accordingly, we obtain the approximate formula

$$\text{Var}(\hat{y}) \doteq \bar{X}^{-2}\{\bar{X}^2 \text{ Var}(\bar{y}) + \bar{Y}^2 \text{ Var}(\bar{x}) + 2 \bar{X} \bar{Y} \text{ Cov}(\bar{x}, \bar{y})\} \qquad (5.5.1)$$

which is an approximately unbiased estimator. For unbiased estimates for these terms, we have following

$$E \frac{1}{n} (1 - \frac{n}{N}) s_y^2 = \text{Var } \bar{y} \qquad (5.5.2)$$

$$E \frac{1}{n} (1 - \frac{n}{N}) s_x^2 = \text{Var } \bar{x} \qquad (5.5.3)$$

and approximately

$$E \bar{y}^2 \doteq \bar{Y}^2 \qquad (5.5.4)$$

$$E \bar{X} \bar{y} \frac{1}{n} (1 - \frac{n}{N}) s_{xy} \frac{N-1}{N} \doteq \bar{X} \bar{Y} \text{ Cov } \bar{x} \bar{y} \qquad (5.5.5)$$

Now upon substituting (5.5.2) to (5.5.5) into (5.5.1) we obtain the approximate estimate of $\text{Var}(\hat{y})$.

$$\text{Var}(\hat{y}) = \frac{1}{n} (1 - \frac{n}{N}) \bar{X}^{-2} \{\bar{X}^2 s_{y_i}^{\ 2} + \bar{y}^2 s_{x_i}^{\ 2} + 2 \bar{X} \bar{y} s_{x_i y_i}\} \qquad (5.5.6)$$

This is an approximately unbiased estimate of the variance of the product of means estimator (for $\bar{X}$ known).

$$\hat{y} = \frac{\bar{x} \bar{y}}{\bar{X}} \qquad (5.5.7)$$

6. UNBIASED PRODUCT ESTIMATORS

The necessity for obtaining unbiased product estimators arises in three main situations:

(i) If it is desired to estimate $\overline{Y}$ with the help of a known population mean $\overline{X}$ and a sample of paired observations (y_i, x_i) which are negatively correlated so that the products $p_i = y_i x_i$ have a smaller coefficient of variation than the y_i.

(ii) If it is desired to estimate the product of two means $\overline{X}\,\overline{Y}$ unbiasedly rather than the mean of product $E(x_i y_i) = \overline{P}$. This situation is comparatively rare.

(iii) If it is desired to estimate $P = E(x_i y_i)$ but only a small sample of paired observations (x_i, y_i) is available while additional independent samples are available for the y_i and/or x_i.

This situation frequently arises.

The principle of unbiased estimation is based on the basic formula (3.1.1) which we may write in the form

$$\overline{P} = E(p_i) = \overline{X}\,\overline{Y} + \text{Cov}(x_i, y_i) \qquad (6.1)$$

or alternatively in the form

$$\overline{Y} = \overline{X}^{-1}\{\overline{P} - \text{Cov}(x_i, y_i)\} \qquad (6.2)$$

In case (i) we employ formula (6.2) and the simple sample of paired observations to estimate $\overline{P}$ and $\mathrm{Cov}(x_i, y_i)$, in case (ii) we employ the same formula and method multiplied by $\overline{X}$ and finally in case (iii) we employ formula (6.1) and use the sample of paired (x_i, y_i) to compute $\overline{p}$ to estimate $\overline{P}$ directly and $\mathrm{Cov}(x_i, y_i)$. An additional reinforcing (independent) estimate of $\overline{X}\,\overline{Y}$ can then be obtained from the independent x and/or y samples. This case is discussed in section 7.

6.1 Unbiased Estimators

We now turn to equation (6.1) and the estimation of the covariance $\mathrm{Cov}(x_i, y_i)$. We have

$$\mathrm{Cov}(x_i, y_i) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{X})(y_i - \overline{Y}) \tag{6.1.1}$$

and estimate this unbiasedly by

$$\widehat{\mathrm{Cov}}(x_i, y_i) = \frac{N-1}{N(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \tag{6.1.2}$$

using the estimate $\overline{p}$ for $\overline{P}$ and substituting (6.1.2) in (6.2) we obtain the unbiased product estimator

$$\hat{\overline{Y}} = \overline{X}^{-1} \{\overline{p} - \frac{N-1}{N(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})\} \tag{6.1.3}$$

which is an unbiased estimator of $\overline{Y}$.

Equation (6.1.3) can be written as

$$\hat{\bar{Y}} = \bar{X}^{-1} \{\bar{p} - \frac{n(N-1)}{N(n-1)} (\bar{p} - \bar{x}\,\bar{y})\} \qquad (6.1.4)$$

6.2 The Exact Variance Formula for $\hat{\bar{Y}}$

In this section we will find the exact variance formula for $\hat{\bar{Y}}$. Recall equation (6.1.4), an unbiased estimator of $\bar{Y}$, which can be written as

$$\hat{\bar{Y}} = \bar{X}^{-1} \{\bar{p} - \frac{N-1}{N} k_{11}\} \qquad (6.2.1)$$

where

$$\frac{n-1}{n} k_{11} = \bar{p} - \bar{y}\,\bar{x} \qquad (6.2.2)$$

and k_{11} is the bivariate k-statistic for the parent bivariate population of $x_i y_i$. We have

$$E(k_{11}) = \kappa_{11} \qquad (6.2.3)$$

where k_{11} is the corresponding k-statistic for the finite population.

Now the variance of (6.2.1) can be written as

$$\mathrm{Var}(\bar{Y}) = \bar{X}^{-2} \{\mathrm{Var}\ \bar{p} - 2(\frac{N-1}{N}) \mathrm{Cov}(\bar{p}, k_{11}) + (\frac{N-1}{N})^2 \mathrm{Var}\ k_{11}\} \qquad (6.2.4)$$

where (see e.g., Kendall and Stuart [8])

$$\mathrm{Var}(k_{11}) = \frac{N-n}{(n-1)(N+1)(N-2)} \left\{ \frac{N-2}{nN} (nN - N - n - 1)\, \kappa_{22} \right.$$

$$\left. + (N - 1)\, \kappa_{20}\kappa_{02} + (N - 3)\, \kappa_{11}^{\;2} \right\} . \qquad (6.2.5)$$

Now the k's can be found from the following equations:

$$k_{11} = \frac{1}{n(n-1)} (n\, S_{11} - \frac{1}{n} S_{10}S_{01}) \qquad (6.2.6)$$

$$k_{21} = \frac{n}{n(n-1)(n-2)} (n^2 S_{21} - 2nS_{10}S_{11} - nS_{20}S_{01} + 2S_{10}^{\;2}\, S_{01}) \quad (6.2.7)$$

and

$$k_{22} = \frac{n}{(n-1)(n-2)(n-3)} \{(n + 1)\, S_{22} - \frac{2(n+1)}{n} S_{21}S_{01}$$

$$- \frac{2(n+1)}{n} S_{12}S_{10} = \frac{n-1}{n} S_{20}S_{02} - \frac{2(n-1)}{n} S_{11}^{\;2} + \frac{8}{n} S_{11}S_{10}S_{01}$$

$$+ \frac{2}{n} S_{02}S_{10}^{\;2} + \frac{2}{n} S_{20}S_{01}^{\;2} - \frac{6}{n^2} S_{10}^{\;2}\, S_{01}\} \qquad (6.2.8)$$

where

$$S_{ab} = \sum_{i=1}^{n} x_i^a\, y_i^b$$

The corresponding values for the κ's are obtained by replacing n by N. Recalling that

$$\mathrm{Var}(\bar{p}) = \frac{S_p^2}{n}\left(1 - \frac{n}{N}\right) \tag{6.2.9}$$

we need only to find the $\mathrm{Cov}(\bar{p}, \kappa_{11})$. To find the $\mathrm{Cov}(\bar{p}, \kappa_{11})$ we write

$$\bar{y}\,\bar{x} = \bar{p} - \frac{n-1}{n}\kappa_{11} \tag{6.2.10}$$

then

$$\mathrm{Var}(\bar{y}\,\bar{x}) = \mathrm{Var}(\bar{p}) - 2\,\frac{n-1}{n}\,\mathrm{Cov}(\bar{p}, \kappa_{11}) + \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\,\kappa_{11} \tag{6.2.11}$$

From (6.2.11) we can write

$$\mathrm{Cov}(\bar{p}, \kappa_{11}) = \frac{n}{2(n-1)}\quad \mathrm{var}(\bar{p}) + \left(\frac{n-1}{n}\right)^2 \mathrm{var}\,\kappa_{11} - \mathrm{var}(\bar{y}\,\bar{x})\} \tag{6.2.12}$$

where from our fundamental product formula (3.1.3) we have

$$\begin{aligned} \mathrm{var}(\bar{y}\,\bar{x}) = {} & \bar{X}^2\,\mathrm{var}\,\bar{y} + \bar{Y}^2\,\mathrm{var}\,\bar{x} + E(\varepsilon_{\bar{x}}^2\,\varepsilon_{\bar{y}}^2) \\ & - \mathrm{cov}^2(\bar{x}, \bar{y}) + 2\,\bar{Y}\,E(\varepsilon_{\bar{x}}^2\,\varepsilon_{\bar{y}}) + 2\,\bar{X}\,E(\varepsilon_{\bar{x}}\,\varepsilon_{\bar{y}}^2) \\ & + 2\,\bar{X}\,\bar{Y}\,\mathrm{cov}(\bar{x}, \bar{y}) \end{aligned} \tag{6.2.13}$$

Now, substituting (6.2.5), (6.2.9) and (6.2.12) into (6.2.4) we can write

$$\begin{aligned}\mathrm{Var}(\hat{\bar{y}}) = \bar{X}^{-2} \{&\mathrm{var}\ \bar{p} - 2(\frac{n}{2n-2})(\frac{N-1}{N})\ \mathrm{var}\ \bar{p} \\ &- 2\ \frac{n-1}{n}\ \frac{N-1}{n}\ \frac{n}{2n-2}\ \frac{N-n}{(n-1)(N+1)(N-2)}\ [\frac{N-2}{nN}(nN - N - n - 1)\kappa_{22} \\ &+ (N-1)\kappa_{20}\kappa_{02} + (N-3)\kappa_{11}{}^2] \\ &+ 2\ \frac{N-1}{N}\ \frac{n}{2n-2}\ [\bar{X}^2\ \mathrm{var}\ \bar{y} + \bar{Y}^2\ \mathrm{var}\ \bar{x} + E(\varepsilon_{\bar{x}}^2\ \varepsilon_{\bar{y}}^2) \\ &- \mathrm{cov}^2(\bar{x}, \bar{y}) + 2\ \bar{Y}\ E(\varepsilon_{\bar{x}}^2\ \varepsilon_{\bar{y}}) + 2\ \bar{X}\ E(\varepsilon_{\bar{x}}\ \varepsilon_{\bar{y}}^2) \\ &+ 2\ \bar{X}\ \bar{Y}\ \mathrm{cov}(\bar{x}, \bar{y})] \\ &+ (\frac{N-1}{N})^2\ \frac{N-n}{(n-1)(N+1)(N-2)}\ \frac{N-2}{nN}(nN - N - n - 1)\kappa_{22} \\ &+ (\frac{N-1}{N})^2\ \frac{N-n}{(n-1)(n+1)(N-2)}\ (N-3)\kappa_{11}{}^2\end{aligned} \qquad (6.2.14)$$

This is the exact formula for $\mathrm{Var}(\hat{\bar{y}})$. Using the theory of k-statistics in finite populations we could estimate (although with considerable effort) the above terms unbiasedly (recognizing such relations as $Ek_{20}k_{02} = Ek_{20}Ek_{02} + \mathrm{Cov}(k_{20}k_{02})$).

We confine ourselves here to estimating approximately unbiasedly the leading terms of (6.2.14) namely var $\bar{p}$, $\bar{X}^2$ var $\bar{y}$, $\bar{Y}^2$ var $\bar{x}$

and $2\ \overline{X}\ \overline{Y}\ \mathrm{cov}(\overline{x}, \overline{y})$. This yields the following approximate formula for the variance estimate of the unbiased product estimator.

$$\mathrm{Var}(\hat{\overline{y}}) \doteq \overline{X}^2 \widehat{\mathrm{Var}}\ \overline{p} + 2\ \frac{N-1}{N}\ \frac{n}{2n-2}\ \widehat{\mathrm{Var}}\ \overline{y}$$

$$+ 2\ \frac{N-1}{N}\ \frac{n}{2n-2}\ \overline{X}^{-2}\ \overline{y}^2\ \widehat{\mathrm{Var}}\ \overline{x}$$

$$+ 4\ \frac{N-1}{N}\ \frac{n}{2n-2}\ \overline{X}^{-1}\ \overline{y}\ \widehat{\mathrm{Cov}}\ (\overline{x}, \overline{y}) \qquad (6.2.15)$$

7. PRODUCT ESTIMATORS WITH ONE FACTOR COMPUTED FROM A SUBSAMPLE

A particularly important application of the formulas of section 3 arises when we compute a product of means estimator

$$\hat{p} = \bar{u}_n \cdot \bar{v}_\kappa \tag{7.1}$$

in which the first factor, $\bar{u}$, is computed from a 'large' sample of n units while the second factor, $\bar{v}_\kappa$, is computed from a subsample of κ units. This situation frequently arises when the characteristic y_i are only available for the subsample. Indeed, the example in crop yield estimation discussed in section 8 is of this type.

Let n be the size of the large sample whose mean is $\bar{u}$. Let κ be the size of the subsample whose mean is $\bar{u}_\kappa$. Therefore the sample of 'left-over' units is of size $\nu = n - \kappa$, and we denote their mean by $\bar{u}_\nu$. Then

$$\bar{U}_n = \frac{\nu\bar{u}_\nu + \kappa\bar{u}_\kappa}{n} \tag{7.2}$$

Therefore we may write

$$\bar{U}_n\bar{V}_\kappa = \frac{1}{n}\{\nu\bar{u}_\nu\bar{v}_\kappa + \kappa\bar{u}_\kappa\bar{v}_\kappa\} \tag{7.3}$$

Recall, from (3.1.5), the exact variance formula for $Var(\bar{U}_n\bar{V}_\kappa)$

$$\mathrm{Var}(\bar{u}_n \bar{v}_\kappa) = \bar{U}^2 \, \mathrm{Var}(\bar{v}_\kappa) + \bar{V}^2 \, \mathrm{Var}(\bar{u}_n)$$

$$+ \, E(\varepsilon\bar{u}_n{}^2 \; \varepsilon\bar{v}_\kappa{}^2) - \mathrm{Cov}^2(\bar{u}_n, \bar{v}_\kappa)$$

$$+ \, 2 \, \bar{V} \, E(\varepsilon\bar{u}_n{}^2 \; \varepsilon\bar{v}_\kappa) + 2 \, \bar{U}^2 \, E(\varepsilon\bar{u}_n \; \varepsilon\bar{v}_\kappa{}^2)$$

$$+ \, 2 \, \bar{U} \, V \, \mathrm{Cov}(\bar{u}_n, \bar{v}_\kappa) \qquad (7.4)$$

In order to obtain a formula for $\mathrm{var}(\bar{u}_n\bar{v}_\kappa)$ we may confine ourselves to the three leading terms, namely, $\bar{U}^2 \, \mathrm{Var}(\bar{v}_\kappa)$, $\bar{V}^2 \, \mathrm{Var}(\bar{u}_n)$ and $2 \, \bar{U} \, \bar{V} \, \mathrm{Cov}(\bar{u}_n, \bar{v}_\kappa)$.

Accordingly we obtain the approximate formula

$$\mathrm{Var}(\bar{u}_n \bar{v}_\kappa) \doteq \bar{U}^2 \, \mathrm{Var}(\bar{v}_\kappa) + V^2 \, \mathrm{Var}(\bar{u}_n)$$

$$+ \, 2 \, \bar{U} \, \bar{V} \, \mathrm{Cov}(\bar{u}_n, \bar{v}_\kappa) \qquad (7.5)$$

The justification of confining (7.5) to the leading terms arises from the fact that normally $C.V.(\bar{u}_n)$ and $C.V.(\bar{v}_\kappa)$ are small, and the three terms retained are those of second order in the C.V.'s. While the others are of 3^{rd} or 4^{th} order in the coefficients of variation.

The third retained term, the covariance term, can be shown to be given by

$$\mathrm{Cov}(\bar{u}_n, \bar{v}_\kappa) \doteq \frac{\kappa}{n} \cdot \frac{1}{\kappa} (1 - \frac{\kappa}{n}) S_{uv} - \frac{\nu}{N} S_{uv}$$

$$= \frac{1}{n} (1 - \frac{\kappa}{N}) S_{uv} - \frac{\nu}{n} \frac{1}{N} S_{uv}$$

$$= \frac{1}{n} (1 - \frac{\nu+\kappa}{N}) S_{uv}$$

$$= \frac{1}{n} (1 - \frac{n}{N}) S_{uv} \qquad (7.6)$$

Here

$$S_{uv} = (N - 1)^{-1} \sum_{i=1}^{N} (u_i - \bar{U})(v_i - \bar{V}) \qquad (7.7)$$

Using (7.6) and the familiar formulas for $\mathrm{Var}(\bar{u}_n)$ and $\mathrm{Var}(\bar{v}_\kappa)$ we may rewrite (7.5) in the final form

$$\mathrm{Var}(\bar{u}_n \bar{v}_\kappa) \doteq \bar{U}^2 \frac{1}{\kappa} (1 - \frac{\kappa}{N}) S_v^2 + \bar{V}^2 \frac{1}{n} (1 - \frac{n}{N}) S_u^2$$

$$+ 2 \bar{U} \bar{V} \frac{1}{n} (1 - \frac{n}{N}) S_{uv} \qquad (7.8)$$

An estimate (although not necessarily unbiased) of $\mathrm{Var}(\bar{u}_n \bar{v}_\kappa)$ can be obtained by substituting unbiased estimates of $\bar{U}$, $\bar{V}$, S_u^2, S_v^2, and S_{uv} in (7.8) so that

$$\hat{\text{Var}}(\bar{u}_n \bar{v}_\kappa) = \bar{u}_n^{\,2} \frac{1}{\kappa} (1 - \frac{\kappa}{N}) s_v^{\,2} + \bar{v}_\kappa^{\,2} \frac{1}{n} (1 - \frac{n}{N}) s_u^{\,2}$$

$$+ 2 \bar{u}_n \bar{v}_\kappa \frac{1}{n} (1 - \frac{n}{N}) s_{uv} \qquad (7.9)$$

where

$$s_u^{\,2} = (n - 1)^{-1} \sum_{i=1}^{n} (u_i - \bar{u}_n)^2$$

$$s_v^{\,2} = (\kappa - 1)^{-1} \sum_{i=1}^{\kappa} (v_i - \bar{v}_\kappa)^2$$

$$s_{uv} = (\kappa - 1)^{-1} \sum_{i=1}^{\kappa} (u_i - \bar{u}_\kappa)(v_i - \bar{v}_\kappa) \qquad (7.10)$$

An application of this formula will be considered in section 8.

8. APPLICATION TO CROP ESTIMATION

In this section we apply the concepts of product estimation developed in the preceding section. Specifically, we use the concept of section 7 dealing with the situation in which one factor is only recorded for a subsample of the original units.

The data used for this illustration were provided by the U.S.D.A. Statistical Reporting Service. They represent a combination of their annual "enumerative survey" conducted in June for the purpose of crop estimation in all agricultural states. The data are concerned with corn estimation in the state of Iowa. From the "enumerative survey" there are available (among other statistics)

$$U = \text{the acres of corn planted} \tag{8.1}$$

for each of a number of sample "tracts". Actually the tracts for which U is recorded are "clustered" in "segments" but the cluster aspect of the data is ignored for the purposes of the present illustration. The calculations shown below deal with all 163 tracts sampled in an "enumeration district".

In a special study (not conducted as a regular annual routine) and concerned with direct measurement there were recorded for a subsample of 30 of the above 163 tracts two additional items, namely

$$Z = (\text{proportion of corn acres harvested})/(\text{corn acres planted}) \tag{8.2}$$

as well as

$$W = \text{yield per acre harvested} \tag{8.3}$$

Since for the subsample of 30 tracts the corn production is clearly obtained as a triple product U·Z·W, where

$$\text{Production} = U \cdot Z \cdot W \tag{8.4}$$

no problem would arise in obtaining an estimate of the total production since the triple product could be regarded as a characteristic attached to the sampled tract. The application of product estimators in double sampling now recognizes that the items Z and W are only recorded for a subsample of 30 tracts while the total corn acres planted (U) is available for the original sample of 163 tracts. Accordingly, using the methods developed in section 7, the total corn production for the enumeration district can be estimated by the product of the mean of U estimated from the large sample multiplied by the mean value of the product Z·W denoted by V which is estimated from the subsample using formulas of section 7. It should be noted that the quantity V is given

$$V = Z \cdot W = \text{total yield harvested/corn acres planted} \tag{8.5}$$

and was computed by this formula (8.5). The estimates of the mean corn production per tract is then multiplied by the "expansion factor" applicable to the enumeration district to obtain an estimate of the total corn production for the enumeration district.

The formulas of section 7 apply, of course, to a random sample and as mentioned above the clustering of the tracts within segments is ignored in our calculation. The formulas require the evaluation of the covariance between U and V and is here estimated from the subsample since only it provides records of both factors U and V. The results of our calculations are shown below, by giving first estimates and estimates of variance for a mean per tract production and subsequently estimates for the total production of the enumeration district.

Now, we can write (8.4) as

$$\text{Production} = N\ \overline{UZW} \tag{8.6}$$

which reduces from the product of three to the product of two random variables. Noting that the sample containing V_κ is a subsample of the sample containing U_n. This is exactly the situation which is dealt with in more detail in section 7.

We also obtained an approximate formula for $\text{Var}(\bar{u}_n \bar{v}_\kappa)$:

$$\text{Var}(\bar{u}_n \bar{v}_\kappa) \doteq \bar{U}^2 \frac{1}{n} (1 - \frac{\kappa}{N})\ s_v^{\ 2} + \bar{V}^2 \frac{1}{n} (1 - \frac{n}{N})\ s_u^{\ 2}$$

$$+ 2\ \bar{U}\ \bar{V} \frac{1}{n} (1 - \frac{n}{N})\ s_{uv} \tag{8.7}$$

and the estimate,

$$\mathrm{Var}(\bar{u}_n \bar{v}_\kappa) = \bar{u}_n^{\,2} \frac{1}{\kappa} (1 - \frac{\kappa}{N}) s_v^{\,2} + \bar{v}^{\,2} \frac{1}{n} (1 - \frac{n}{N}) s_u^{\,2}$$

$$+ 2 \bar{u}_n \bar{v}_\kappa \frac{1}{n}(1 - \frac{n}{N}) s_{uv} \qquad (8.8)$$

(not necessarily unbiased) was obtained. Here we have

$$s_u^{\,2} = (n - 1)^{-1} \sum_{i=1}^{n} (u_i - \bar{u}_n)^2 \qquad (8.9)$$

$$s_v^{\,2} = (\kappa - 1)^{-1} \sum_{i=1}^{\kappa} (v_i - \bar{v}_\kappa)^2 \qquad (8.10)$$

and

$$s_{uv} = (\kappa - 1)^{-1} \sum_{i=1}^{\kappa} (u_i - \bar{u}_\kappa)(v_i - \bar{v}_\kappa) \qquad (8.11)$$

These formulas were applied to the data from U.S.D.A. Statistical Reporting Service, yield the following calculations:

$$E(v_\kappa) = 110.8 \qquad (8.12)$$

$$E(u_n) = 48.4 \qquad (8.13)$$

$$\mathrm{Cov}(u_n, v_\kappa) = -240.6979 \qquad (8.14)$$

Total number of tracts/segment = 163 (8.15)

Segment total for corn production = 834889.6023 (8.16)

Estimate of the variance $(\bar{u}_n \bar{v}_\kappa)$ = 304662.6942 (8.17)

using

$$s_{u_n}^2 = 1611.3591 \quad (8.18)$$

and

$$s_{v_\kappa}^2 = 2713.2215 \quad (8.19)$$

The estimates (8.12) to (8.16) certainly are of a reasonable magnitude.

REFERENCES

[1] Cochran, W. G. (1963). Sampling Techniques. John Wiley & Sons, Inc.

[2] Goodman, L. A. and Hartley, H. O. (1958). "The Precision of Unbiased Ratio-Type Estimators," Journal of the American Statistical Association, V. 53, pp. 491-508.

[3] Haldane, J. B. S. (1942). "Moments of the Distribution of Powers and Products of Normal Variates," Biometrika, V. 32, pp. 226-241.

[4] Goodman, L. A. (1960). "On the Exact Variance of Products," Journal of the American Statistical Association, V. 55, No. 292, pp. 708-713.

[5] Murthy, M. N. (1964). "Product Method of Estimation," Sankhya, Series A, 26, pp. 69-74.

[6] Nanjamma, M. S., Murthy, M. N., and Sethi, V. K. (1959), "Some Sampling Systems Providing Unbiased Ratio Estimators." Sankhya, V. 21, pp. 299-314.

[7] Cook, M. B. (1951). "Bivariate k-Statistics." Biometrika, V. 38, p. 179.

[8] Kendall, M. G. and Stuart, A. (1969). The Advanced Theory of Statistics, V. 1. Hafner Publishing Company.

[9] Fisher, R. A. (1929). "Moments and Product Moments of Sampling Distributions," Proceedings of the London Mathematical Society. V. 30, pp. 199-238.

APPENDIX

APPENDIX

General Estimation Formulas

We will now consider estimating unbiasedly the following parameter functions

$$\overline{X}\,\overline{Y},\ \overline{X}^2,\ \overline{Y}^2,\ \mathrm{Cov}(x_i, y_i),\ E(\varepsilon x_i^2 \varepsilon y_i^2),\ E(\varepsilon x_i^2 \varepsilon y_i) \text{ and } E(\varepsilon x_i \varepsilon y_i^2)$$

by using the so-called k-statistics for finite populations. These unbiased estimates will then be substituted in the terms of equation (3.1.5) resulting in approximately unbiased estimates of that variance formula in the following sense. Whenever an additive term consists of a product of two or three of the terms listed above the covariances between the estimates of these terms will be ignored.

An unbiased estimator for the above terms can be found by using the following formulas and defintions and from properties of bivariate k-statistics, (see e.g., Kendall, M.G. and Stuart, A. [8]). They are as follows:

$$\mathrm{Var}(x) = Ex^2 - (Ex)^2 \qquad (A.1)$$

$$\overline{x} = k_{10} \qquad (A.2)$$

$$\overline{y} = k_{01} \qquad (A.3)$$

$$k_{20} = (n - 1)^{-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad (A.4)$$

$$k_{02} = (n - 1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{A.5}$$

and

$$k_{11} = \text{Cov}(x_i, y_i) = (n - 1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{A.6}$$

The above are the sample means, variances and covariance, respectively. Consider now

$$p_i = x_i y_i \tag{A.7}$$

then

$$E\, p_i = \bar{X}\, \bar{Y} + \text{Cov}(x_i, y_i) \tag{A.8}$$

where the $\text{Cov}(x_i, y_i)$ is estimated unbiasedly by

$$\frac{N-1}{N(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{A.9}$$

We can then write

$$\bar{X}\, \bar{Y} = E\{p_i - \frac{N-1}{N(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\}$$

$$= E\{\bar{P} - \frac{N-1}{N(n-1)} (n\, \bar{P} - n\, \bar{x}\, \bar{y})\} \tag{A.10}$$

Therefore $\bar{X}\, \bar{Y}$ is estimated unbiasedly by

$$\widehat{\bar{X}\, \bar{Y}} = \bar{P} - \frac{n(N-1)}{N(n-1)} (\bar{P} - \bar{x}\, \bar{y}) \tag{A.11}$$

Recall now that

$$\mathrm{Var}(\bar{x}) = E\,\bar{x}^2 - [E\,\bar{x}]^2 \tag{A.12}$$

and the corresponding formula for $\mathrm{Var}(\bar{y})$. Therefore unbiased estimators of $\bar{X}^2$ and $\bar{Y}^2$ are given by

$$\hat{\bar{X}}^2 = \bar{x}^2 - \frac{N-n}{nN} s_x^2 \tag{A.13}$$

and

$$\hat{\bar{Y}}^2 = \bar{y}^2 - \frac{N-n}{nN} s_y^2 \tag{A.14}$$

respectively. Now to find $E\varepsilon x_i^2 \varepsilon y_i^2$, $E\varepsilon x_i^2 \varepsilon y_i$ and $E\varepsilon x_i \varepsilon y_i$ we will use k-statistics. We consider the family of k-statistics k_{11}, k_{12}, ..., k_{pq}, ... which are functions of the observations and are such that the mean value of k_{pq} is the pq^{th} cumulant K_{pq}, i.e.,

$$E\,k_{pq} = K_{pq} \tag{A.15}$$

The following is a relation of a function for a sample to the corresponding function for the parent. Denoting the two functions by the subscripts n and N respectively, and if K_{pq} is the pq^{th} k-statistic for the parent, we have (see e.g., Kendall and Stuart [8]).

$$E_N\,k_{pq} = K_{pq} \tag{A.16}$$

where E_N means an expectation in the set of N values forming the parent. This attractive property of k-statistics is one of their most useful features. Now to estimate $E_N \varepsilon x_i \varepsilon y_i = N^{-1}\Sigma(x_i - \bar{X})(y_i - \bar{Y})$ unbiasedly we consider the following:

$$E_N\, k_{11} = K_{11} \tag{A.17}$$

where k_{11} for an infinite population is given by

$$k_{11} = \frac{1}{n(n-1)}\,(n\, S_{11} - S_{10}\, S_{01}) \tag{A.18}$$

(see e.g., Kendall, M.G. and Stuart, A. [8]). Where

$$S_{pq} = \Sigma x_i^p\, y_i^q \tag{A.19}$$

now by noting that

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - \frac{1}{n}\Sigma x_i\, \Sigma y_i - \frac{1}{n}\Sigma x_i\, \Sigma y_i$$

$$+ \frac{n}{n^2}\Sigma x_i\, \Sigma y_i$$

$$= S_{11} - \frac{1}{n} S_{01}\, S_{10}$$

We can write

$$\frac{n}{n(n-1)}\sum_i\,(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n(n-1)}\,[n\, S_{11} - S_{01}\, S_{10}]$$

therefore

$$K_{11} = \frac{N}{N(N-1)} \sum_i (x_i - \bar{X})(y_i - \bar{Y}) \tag{A.20}$$

and (for finite population)

$$k_{11} = \frac{n}{n(n-1)} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \tag{A.21}$$

i.e., sampling from a finite population

$$\hat{E}(\varepsilon x_i \varepsilon y_i) = \frac{N}{n(N-1)} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \tag{A.22}$$

Now to estimate $E(\varepsilon x_i^2 \varepsilon y_i)$ unbiasedly we use the relation

$$E_N \, k_{21} = K_{21} \tag{A.23}$$

where k_{21} for an infinite population is given by

$$k_{21} = \frac{1}{n(n-1)(n-2)} [n^2 S_{21} - 2nS_{10} \, S_{11} - nS_{20} \, S_{01} + 2S_{10}^2 \, S_{01}] \tag{A.24}$$

(see e.g., Kendall, M. G. and Stuart, A. [8]). Where S_{pq} is given by (A.19). Now by noting that

$$\Sigma(x_i - \bar{x})^2 \, (y_i - \bar{y}) = \Sigma x_i^2 \, y_i - 2\bar{x} \, \Sigma x_i y_i + \bar{x}^2 \, \Sigma y_i - \bar{y} \, \Sigma x_i^2$$

$$+ 2 \, \bar{x} \, \bar{y} \, \Sigma x_i - n \, \bar{x}^2 \, \bar{y}$$

$$= S_{21} - \frac{2}{n} S_{10} \, S_{11} - \frac{1}{n} S_{01} \, S_{20} + \frac{2}{n^2} S_{10}^2 \, S_{01}$$

We can write

$$\frac{n}{(n-1)(n-2)} \sum_i (x_i - \bar{x})(y_i - \bar{y}) = k_{21} \tag{A.25}$$

therefore

$$K_{21} = \frac{N}{(N-1)(N-2)} \sum_i^N (x_i - \bar{X})(y_i - \bar{Y}) \tag{A.26}$$

and (for a sample of n from the finite population)

$$k_{21} = \frac{n}{(n-1)(n-2)} \sum_i^n (x_i - \bar{x})(y_i - \bar{y}) \tag{A.27}$$

i.e., sampling from a finite population.

$$\hat{E}\ \varepsilon x_i^2\ \varepsilon y_i = \frac{(N-1)(N-2)}{N^2} \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y}) \tag{A.28}$$

Likewise

$$\hat{E}\ \varepsilon x_i\ \varepsilon y_i^2 = \frac{(N-1)(N-2)\ n}{N^2(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})^2 \tag{A.29}$$

Now to estimate $E(\varepsilon x_i^2\ \varepsilon y_i^2)$ unbiasedly we use the relation

$$E_N\ k_{22} = K_{22} \tag{A.30}$$

where k_{22} for both a finite and an infinite population is given by

$$k_{22} = \frac{n}{(n-1)(n-2)(n-3)} \{(n+1)S_{22} - \frac{2(n+1)}{n} S_{21}S_{01} - \frac{2(n+1)}{n} S_{12}S_{10}$$

$$- \frac{(n-1)}{n} S_{20}S_{02} - \frac{2(n-1)}{n} S_{11}^2 + \frac{8}{n} S_{11}S_{10}S_{01} + \frac{2}{n} S_{02}S_{10}^2$$

$$+ \frac{2}{n} S_{20}S_{01}^2 - \frac{6}{n^2} S_{10}^2 S_{01}^2 \qquad \text{(A.31)}$$

(see e.g., Kendall, M. G. and Stuart, A. [8]) where S_{pq} is given by (A.19). Now we note that

$$\Sigma(x_i - \bar{x})^2 (y_i - \bar{y})^2 = S_{22} - \frac{2}{n} S_{01}S_{21} + \frac{1}{n^2} S_{01}^2 S_{20} - \frac{2}{n} S_{10}S_{12}$$

$$+ \frac{4}{n^2} S_{10}S_{01}S_{11} - \frac{2}{n^3} S_{10}^2 S_{01}^2 + \frac{1}{n^2} S_{10}^2 S_{02}$$

$$- \frac{2}{n^3} S_{10}^2 S_{01}^2 + \frac{1}{n^3} S_{10}^2 S_{01}^1 \qquad \text{(A.32)}$$

and our task is to add terms to (A.32) so that the result will yield (A.31). To do this we will first add the term $\{- \frac{2}{n} k_{11}^2 (n-1)^3\}$, where

$$k_{11} = \frac{1}{n(n-1)} (nS_{11} - S_{10}S_{01}) \qquad \text{(A.33)}$$

(see e.g., Kendall, M. G. and Stuart, A. [8]) where S_{pq} is given by (A.19). Now we can write

$$k_{11}^2 = \frac{1}{n^2(n-1)^2} (n^2 S_{11}^2 - 2nS_{11}S_{10}S_{01} + S_{10}^2 S_{01}^2)$$

and

$$-\frac{2}{n}(n-1)^3 k_{11}^2 = \frac{4}{n^2}(n-1)\ S_{11}S_{10}S_{01} - \frac{2}{n^3}(n-1)\ S_{10}^2 S_{01}^2$$

$$-\frac{2(n-1)}{n} S_{11}^2 \qquad \text{(A.34)}$$

Now we consider the term

$$T = (n-1)^2\ k_{10}^2 k_{02} + 2n(n-1)\ k_{01}^2 k_{10}^2 + (n-1)^2 k_{01}^2 k_{20} \qquad \text{(A.35)}$$

where

$$k_{10} = \frac{1}{n} S_{10} \qquad \text{(A.36)}$$

$$k_{20} = \frac{1}{n-1}(S_{20} - \frac{1}{n} S_{10}^2) \qquad \text{(A.37)}$$

$$k_{01} = \frac{1}{n} S_{01} \qquad \text{(A.38)}$$

$$k_{02} = \frac{1}{n-1}(S_{02} - \frac{1}{n} S_{01}^2) \qquad \text{(A.39)}$$

Now we can write

$$k_{10}^2 = \frac{1}{n^2} S_{10}^2 \qquad \text{(A.40)}$$

$$k_{01}^2 = \frac{1}{n^2} S_{01}^2 \qquad \text{(A.41)}$$

Now substituting (A.37), (A.39), (A.40) and (A.41) into (A.35), we can then write (A.35) as

$$T = \frac{n-1}{n^2} s_{10}^2 s_{02} + \frac{n-1}{n^2} s_{01}^2 s_{20} - \frac{(n-1)^2}{n^3} s_{10}^2 s_{01}^2 - \frac{(n-1)^2}{n^3} s_{01}^2 s_{10}^2$$

$$+ \frac{2}{n^3} (n-1)\ s_{01}^2 s_{10}^2 \tag{A.42}$$

and the last term added is

$$\tau = -\ n(5n+1)\ k_{01}^2 k_{10}^2 \tag{A.43}$$

where k_{10}^2 and k_{01}^2 are given by (A.40) and (A.41) respectively. Now substituting (A.40) and (A.41) into (A.43) we can write

$$\tau = -\ \left(\frac{5}{n^2} s_{01}^2 s_{10}^2 + \frac{1}{n^3} s_{01}^2 s_{10}^2 \right) \tag{A.44}$$

Now combining terms in (A.32), (A.34), (A.42) and (A.44) we can write the term

$$(n+1)\ s_{22} - \frac{2}{n} (n+1)\ s_{01} s_{21} + \frac{2}{n} s_{01}^2 s_{20} - \frac{2}{n} (n+1)\ s_{10} s_{12}$$

$$+ \frac{8}{n} s_{10} s_{01} s_{11} + \frac{2}{n} s_{02} s_{10}^2 - \frac{2}{n} (n-1)\ s_{11}^2 - \frac{4}{n^2} s_{10}^2 s_{01}^2$$

$$- \frac{6}{n^3} s_{10}^2 s_{01}^2 + \frac{2}{n} s_{10}^2 s_{01}^2 \tag{A.45}$$

In term (A.45) we need the term $- \frac{n-1}{n} s_{20} s_{02}$. Therefore we consider adding the term $\frac{-(n-1)(n-2)(n-3)}{n} k_{22}$ to (A.45) thus finally we can write

$$k_{22} = \frac{n}{(n-1)(n-2)(n-3)} \{(n+1)\ \Sigma(x_i - \bar{x})^2\ (y - \bar{y})^2 - \frac{2}{n}\ (n-1)^3 k_{11}^2$$

$$+\ 2(n-1)^2\ k_{10}^2 k_{02} + 3n(n-1)\ k_{01}^2 k_{10}^2 + 2(n-1)^2\ k_{01}^2 k_{20}$$

$$-\ \frac{(n-1)(n-2)(n-3)}{n}\ k_{22}\} \qquad \text{(A.46)}$$

From (A.46) it follows that

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)}\ \Sigma(x_i - \bar{x})^2 (y_i - \bar{y})^2 = k_{22} + \frac{2(n-1)^2}{(n-2)(n-3)}\ k_{11}^2$$

$$-\ \frac{2n(n-1)}{(n-2)(n-3)}\ k_{10}^2 k_{02} - \frac{3n^2}{(n-2)(n-3)}\ k_{01}^2 k_{10}^2$$

$$-\ \frac{2n(n-1)}{(n-2)(n-3)}\ k_{01}^2 k_{20} \qquad \text{(A.47)}$$

The corresponding formula for finite populations is given by

$$K_{22} = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y}) - \frac{2(N-1)^2}{(N-2)(N-3)}\ K_{11}^2$$

$$+\ \frac{2N(N-1)}{(N-2)(N-3)}\ K_{10}^2 K_{02} + \frac{3N^2}{(N-2)(N-3)}\ K_{01}^2 K_{10}^2$$

$$+\ \frac{2N(N-1)}{(N-2)(N-3)}\ K_{01}^2 K_{20} \qquad \text{(A.48)}$$

Expressing (A.48) as an equation for $E\ \varepsilon x_i^2\ \varepsilon y_i^2 = N^{-1}\Sigma\ (x_i-\bar{X})^2 (y_i-\bar{Y})^2$ we obtain

$$E\ \varepsilon x_i^2\ \varepsilon y_i^2 = \frac{(N-1)(N-2)(N-3)}{N^2(N+1)} K_{22} + \frac{2(N-1)^3}{N^2(N+1)} K_{11}^2$$

$$- 2\,\frac{(N-1)^2}{N(N+1)}\ (K_{10}^2 K_{02} + K_{01}^2 K_{20})$$

$$- \frac{3(N-1)}{(N+1)} K_{01}^2 K_{10}^2 \qquad \text{(A.49)}$$

In order to estimate $E\ \varepsilon x_i^2\ \varepsilon y_i^2$ we estimate unbiassedly K_{22} by k_{22}; K_{02} by k_{02}; K_{20} by k_{20}; K_{10}^2 by $k_{10}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{20}$ and K_{01}^2 by $k_{01}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{02}$. For the remaining terms we only employ the leading terms ignoring variances and covariances. That is we substitute in (A.49) k_{11}^2 for K_{11}^2, $(k_{10}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{20})k_{02}$ for $K_{10}^2 K_{02}$ $(k_{01}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{02})k_{20}$ for $K_{01}^2 K_{20}$ and $(k_{10}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{20})(k_{01}^2 - \frac{1}{n}(1 - \frac{n}{N})k_{02})$ for $K_{01}^2 K_{10}^2$.