COMPARISON OF ESTIMATORS FOR TWO MULTI-STAGE DESIGNS WHEN SAMPLING ON SUCCESSIVE OCCASIONS

Park Hong Nai

Institute of Statistics
Mimeograph Series No. 503
December 1966

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

1. INTRODUCTION

1.1. Basis for the Present Investigation

In practice, there are many cases where sampling surveys are carried out over time for the purpose of estimating certain population characteristics which vary with time. In such sampling surveys carried out on a series of successive occasions in time, a problem arises, different from that of single occasion surveys, which is concerned with how to utilize the past information to improve the current estimate of a population characteristic of interest.

Specifically, the problem of utilization of past information consists of two parts: one is the scheme for partial retention of units in the sample after each occasion and the other is an estimation problem.

Many studies concerned with this problem have been done. Most of these studies are devoted to

1. sampling with replacement of units at each draw,
2. a partial retention scheme which is repetitive but not periodic,
3. an estimation theory emphasizing a composite estimator or a regression type estimator.

However, regarding these past studies, the following points have not received adequate attention:

1. sampling is usually carried out without replacement at each draw,
2. in a sampling design problem, it is necessary to investigate the effect of the partial retention scheme whereby the units are repetitive but not periodic,

3. in an estimation problem, it is necessary to study the efficiency of the estimators which have been used, and

4. in order to increase the precision of an estimator in multistage sampling, it is preferred that partial retention is made of the first-stage units rather than the second-stage or succeeding stage-units.

This thesis will study the sampling problem on successive occasions for estimating the current population mean where partial retention is made of the first-stage units and the units are drawn without replacement at each draw and with equal probabilities.

1.2. Nature of the Problem

The problem will be formulated as follows:

1. This thesis will treat the sampling problem on successive occasions where a sample is drawn without replacement at each draw. Therefore, in Chapter 3, variances and covariances for subsets on any occasion are derived according to the proposed sampling scheme of partial retention of units.

2. The sampling theory for two successive occasions provides a basis for the theory appropriate to more than two occasions. Therefore, in Chapter 4, the properties of four existing estimators (referred to as a general linear estimator, a modified composite estimator, a composite estimator and a ratio type estimator) and the comparison of their efficiencies are examined for the two successive occasion case. In addition, to the four estimators, a symmetric estimator will be introduced and its efficiency will be examined relative to the other four estimators.

3. The estimation theory obtained for two successive occasions under the proposed sampling scheme will then be extended for more than two successive occasions. In sampling for more than two occasions, we may consider a sequence of estimators of the current population mean corresponding to the number, i say, of previous occasions for which data are included. In Chapter 5, the properties of two selected estimators from Chapter 4 and the determination of the preferred number of occasions i will be examined.

4. The idea of a partial retention scheme is that a part of the sample consisting of Pn units on occasion α is retained in the sample on the next occasion $\alpha + 1$, and the remaining part of the sample is replaced on occasion $\alpha + 1$ by new units distinct from any previous ones to improve the precision of an estimator. Hence, in this context, it is possible to formulate a partial retention pattern where the same Pn units in the first occasion sample are retained on all succeeding occasions and the remaining Qn units $(P + Q = 1)$ are replaced on every occasion. Based on this new partial retention scheme, it may be possible to construct an estimator which has smaller variance than other existing estimators. The efficiency of this new estimator, called a "complete" estimator, and its comparison with other estimators is examined in Chapter 6.

Assumptions which are made in this study are:

1. the population units are fixed on all occasions,
2. the sample size n is constant on all occasions, and
3. the discarded fraction Q is constant on all occasions.

In the theoretical development no assumptions are made regarding the correlation pattern between the same unit values on different occasions. A specific correlation pattern, due to Yates, was assumed for the emperical studies of the relative efficiency of alternative estimators.

2. REVIEW OF LITERATURE

The first study of sampling on successive occasions seems to have been done by Jessen in 1942 [7]. For that investigation, two sample surveys of Iowa farms were made on two successive years. The first survey was done in 1938. The second survey in 1936 was an integral part of the first survey in that 50 percent of the first sample was selected for re-enumeration and the remaining 50 percent was replaced by a new sample. With this design, Jessen obtained two independent estimators of the mean on the second occasion; one was the sample mean based only on the new portion of the second sample, and the other was a regression estimate based on the matched portions and the overall sample mean of the first occasion. A combined estimator of the second occasion mean also was obtained by weighing the two independent estimates inversely as their variances. This was the first exploitation of sampling on successive occasions.

A theory for sampling on more than two occasions was developed by Yates in 1949 [14]. The estimator suggested by Yates is very similar to Jessen's estimator. The only change is to use the weighted estimate for the previous occasion instead of the single occasion sample mean in the regression adjustment of the estimate based on the matched portion. Yates assumes that a given fraction of the units is replaced on each occasion, that the variability on the different occasions and the correlation between successive occasions are constant, and that the correlation between units two occasions apart is r^2 , that between units three occasions apart is r^3, etc.

In 1950, the theoretical development on the estimator given by Yates was completed by Patterson [10]. He formulated the necessary and sufficient conditions for an estimator of the current mean having a specific recursive linear form of variates to be efficient. He also derived the optimum weight for his linear estimator, the limiting value of the weight and some other properties of his estimator.

In 1954, the Bureau of the Census (Hansen, et al.) [5] introduced a redesign of the Current Population Survey, in which one primary unit was selected from a stratum with probabilities proportionate to the size. The subsampling of the selected primary unit involved a scheme of partial replacement or rotation of units at the last stage, in order to avoid a decline in respondent cooperation and to reduce the variances of sample estimates. A new estimator, the so called "composite estimate" was introduced. It is a composite of two estimates. The first is the regular ratio estimate based on the entire sample for month h. The second estimate consists of the estimate for the preceding month plus an estimate of the change from the preceding month to the present month. Under such a sampling scheme, only the within first stage unit component of variance of the estimate is improved while the between first stage unit variance still remains the same as in the regular estimate. It should be pointed out that this composite estimator can be identical with the Yates estimator in many practical situations where the correlation between successive occasions is fairly high. Another linear estimator, the so called "general linear estimator", was given for the infinite population approach for two occasions in [5]. In 1955, Tikkiwal [13] developed the sampling theory for k characters on each occasion from a finite populations, assuming a correlation pattern slightly different from that of Patterson. The previous studies of sampling on successive occasions have been limited to single character estimators on each occasion.

Onate [9] in 1960, in developing multistage sampling designs for the Phillipine Statistical Survey of Households, adopted the same principle as the U. S. Current Population Survey redesign. Moreover, Onate developed a finite population theory for the composite estimator defined in the redesign.

In 1964, Rao and Graham [12] further extended Onate's finite population theory for the composite estimator to a specific rotation pattern which is different from that treated in this thesis as discussed in Chapter 3. The Rao and Graham rotation scheme applies when $\frac{1}{Q}$ is an integer, and also assumes that a unit can return to the sample after having been dropped from the sample. In 1965, Des Raj [2] proposed the selection of clusters with probabilities proportional to size for sampling on two successive occasions and indicated the application of the theory to double sampling.

Purakam in 1966 [11] studied multi-stage sampling on successive occasions. The sampling scheme consists of the selection of the first-stage units with unequal probabilities with replacement at each draw and the selection of the second stage units with equal probabilities but without replacement. A partial replacement of first-stage units is proposed, which is the same as that in this thesis. He considered four types of estimators for sampling on two successive occasions, and also extended the estimation procedure to three successive occasions. These four types of estimators are studied in this work.

From the review of the literature, it can be seen that the "general linear estimator" has still not been investigated for more than two occasion sampling and may be more efficient; that the finite theory for a composite estimator as given by Rao seems to be semi-finite in the sense that the number of previous occasions is assumed to be infinite; and that in most cases the correlation patterns are specified.

3. SAMPLING PROCEDURE AND BASIC RESULTS FOR ESTIMATION THEORY

3.1. The Proposed Sampling Scheme

The following sampling scheme will be adopted throughout this thesis except in Chapter 6.

Assume a population U consisting of N definable first-stage units on α occasions: $\{u_i\}$, $i = 1, 2, \ldots N$. Each u_i contains N_i second-stage units. The sampling scheme is defined as follows:

1. Assume that a constant proportion Q $(0 < Q < 1)$ of first-stage units is to be replaced after each occasion, a preliminary first-stage sample of size $n + (\alpha-1)Qn$ is selected from U with equal probability without replacement of unit at each draw where N is larger than $n + (\alpha-1)Qn$ for any α. The order of the draw for each unit is recorded.

2. The first-stage units which occur from order 1 to n constitute the sample for the first occasion. Then n_i second stage units are drawn from the ith first-stage unit with equal probability and without replacement. This second-stage sampling scheme is applied on every occasion.

3. The first Qn units are rejected and the next $(1-Q)n$ units are retained for the second occasion. The retained units are supplemented by the next set of Qn units which occurred from order $n + 1$ to $n + Qn$. Thus, the required sample size of n first-stage units is maintained on the second occasion with the assurance of having $(1-Q)n$ units matched with those of the first occasion. The subsampling from the Qn first-stage units is done the same as on the first occasion. The second-stage units originally selected from the matched Pn first-stage units are also retained for the

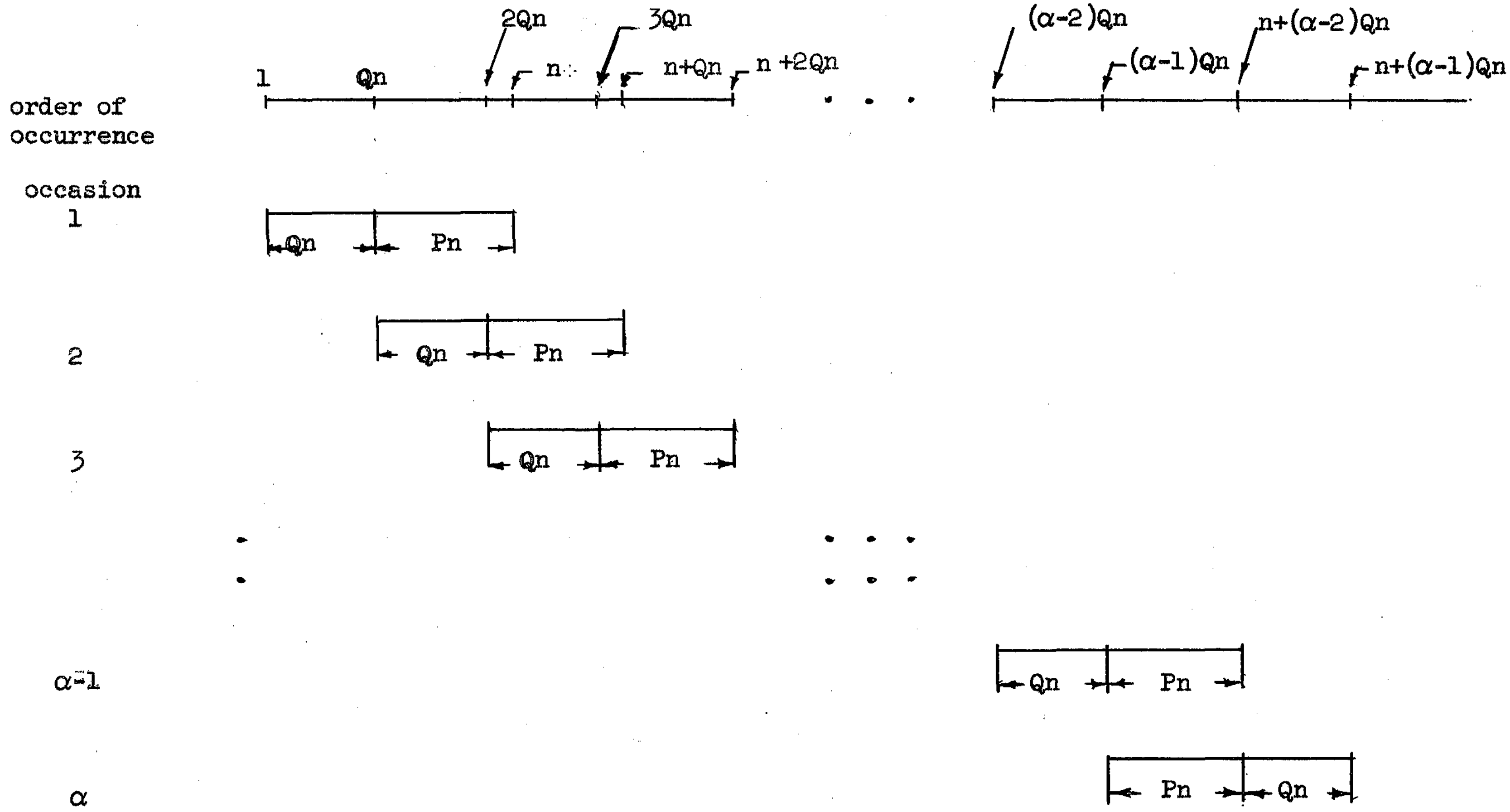


Figure 3.1. Proposed sampling scheme

second occasion. On the third occasion, the next Qn units which occurred from order Qn+1 to 2Qn are rejected while the other (1-Q)n units which occurred from order 2Qn+1 to n+Qn are retained. The retained units are supplemented by the Qn units which occurred from order n+Qn+1 to n+2Qn. This procedure is carried out in a similar fashion for the succeeding occasions. On the αth occasion, there will be (1-Q)n units which are matched with the α-1th occasion plus Qn unmatched units which occurred from order $n + (\alpha-2)Qn+1$ to $n+(\alpha-1)Qn$. Figure 1 may help to understand the structure of the sampling scheme.

3.2. Derivation of Variances and Covariances for Subsets of the Preliminary Sample in Uni-stage Sampling

In developing an estimation theory for the proposed sampling design, the principle of minium variance unbiased estimation will be adopted. Since we will be considering several classes of estimators having a linear form of sample means, it is necessary before entering into the estimation problems to derive general variances and covariances of means on any occasion. The notation to be used is the following:

N = the number of first-stage units in the population on every occasion.

n = the number of first-stage units in the sample on every occasion.

N_i = the number of second-stage units in the ith first-stage unit.

n_i = the number of second-stage units in the sample in the ith first-stage unit.

$m = \sum_{i=1}^{N} n_i$ = the number of the second-stage units in the sample

$Y_{\alpha ij}$ = the value of the jth second-stage unit in the ith first-stage unit on the α occasion.

$Y_{\alpha i} = \sum_{j=1}^{N_i} Y_{\alpha ij}$ = the value of the ith first-stage unit on the α occasion.

$\bar{\bar{Y}}_{\alpha} = \frac{\sum_{i=1}^{N} Y_{\alpha i}}{N}$ = population mean on occasion α.

$\bar{\bar{Y}}_{\alpha i} = \frac{\sum_{j=1}^{N_i} Y_{\alpha ij}}{N_i}$ = mean of the N_i second stage units in the ith first-stage unit on occasion α.

The subscript i will also be used to denote the order of the first-stage units occurring in the various sub-samples according to the sampling scheme. Thus $\bar{Y}_{(\alpha-1)1} = \frac{\Sigma Y_{(\alpha-1)i}}{Qn}$ = sample mean on occasion $(\alpha-1)$ of the Qn first-stage units which were selected in order $i = (\alpha-2)Qn+1, \ldots, (\alpha-1)Qn$.

$\bar{Y}_{(\alpha-1)2} = \frac{\Sigma Y_{(\alpha-1)i}}{Pn}$ = sample mean on occasion $(\alpha-1)$ of the Pn first-stage units which were selected in order $i = (\alpha-1)Qn+1, \ldots, n+(\alpha-2)Qn$.

$\bar{Y}_{\alpha 1} = \frac{\Sigma Y_{\alpha i}}{Pn}$ = sample mean on occasion α of the Pn first-stage units which occurred in order $i = (\alpha-1)Qn+1, \ldots, n+(\alpha-2)Qn$.

$\bar{Y}_{\alpha 2} = \frac{\Sigma Y_{\alpha i}}{Qn}$ = sample mean on occasion α of the Qn first-stage units which were selected in order $i = n+(\alpha-2)Qn+1, \ldots, n+(\alpha-1)Qn$.

$\sigma^2_{\alpha 1} = \frac{1}{N} \sum_{i=1}^{N} (Y_{\alpha i} - \bar{\bar{Y}}_{\alpha})^2$ = variance among first-stage units on occasion α.

$\sigma^2_{\alpha 2i} = \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{\alpha ij} - \bar{\bar{Y}}_{\alpha i})^2$ = variance of second stage units within the ith first-stage unit on occasion α.

$\sigma_{(\alpha-1,\alpha)1} = \frac{1}{N} \sum_{i=1}^{N} (Y_{(\alpha-1)i} - \bar{\bar{Y}}_{(\alpha-1)})(Y_{\alpha i} - \bar{\bar{Y}}_{\alpha})$ = covariance between first-stage unit values on the two successive occasions $(\alpha-1)$ and α.

$\sigma_{(\alpha-1,\alpha)2i} = \frac{1}{N_i} \sum_{j=1}^{N_i} (Y_{(\alpha-1)ij} - \bar{\bar{Y}}_{(\alpha-1)i})(Y_{\alpha ij} - \bar{\bar{Y}}_{\alpha i})$ = covariance between second-stage unit values (within the ith first-stage unit) on the two successive occasions $(\alpha-1)$ and α.

In order to derive variances and covariances for subsets on any occasions, the following two approaches can be considered:

1. By the definition of the preliminary sample as in 3.1, any sub-sample of the preliminary sample is a random sample selected from the population of N units. This approach does not involve the notion of conditional probability. Hence, the following lemmas 1, 2, 3, 6, 7 and 8 follow directly from the appropriate theorems for random samples selected without replacement on a single occasion.

2. Since the sampling is carried out without replacement of units at each draw, the sample on the αth occasion is a random sample selected from the population of $N-(\alpha-1)Qn$ units where $(\alpha-1)Qn$ units have been discarded on $\alpha-1$ previous occasions. This approach involves the notion of conditional probability.

In dealing with the sampling problem on successive occasions, it seems that the second approach is more natural than the first approach. Lemmas 4 and 9 will be derived by the second approach.

Lemma 1. The sample mean of the set of first-stage units which occurred from order $(\alpha-1)Qn+1$ to $n+(\alpha-2)Qn$ in the preliminary sample is an unbiased estimator of $\bar{\bar{Y}}_\alpha$, that is

$$E(\bar{Y}_{\alpha 1} = \frac{\Sigma Y_{\alpha i}}{Pn}) = \bar{\bar{Y}}_\alpha \quad \text{for all } \alpha ,$$

$$i = (\alpha-1)Qn+1, \ldots, n+(\alpha-2)Qn.$$

Lemma 2. The variance of the sample mean $\bar{Y}_{\alpha 1}$ of size Pn on occasion α is

$$V(\bar{Y}_{\alpha 1}) = \frac{\sigma^2_{\alpha 1}}{Pn} \left(\frac{N-Pn}{N-1}\right) \quad \text{for all } \alpha .$$

Lemma 3.

$$\text{Cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 1}) = \frac{\sigma_{(\alpha-1,\alpha)1}}{Pn} \left(\frac{N-Pn}{N-1}\right) \quad \text{for all } \alpha.$$

Lemma 4.

$$\text{Cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 2}) = -\frac{\sigma_{(\alpha-1,\alpha)1}}{N-1} \quad \text{for all } \alpha.$$

Proof:

$$\text{Cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 2}) = \underset{s}{E} \text{ cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 2}/s) + \text{cov } E(\bar{Y}_{(\alpha-1)\alpha}/s),$$

$$E(\bar{Y}_{\alpha 2}/s) \quad \text{where} \quad s = \{u_i : i = 1, \ldots, (\alpha-1)Qn\} .$$

Consider $E \text{ cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 2}/s)$.

$$\text{cov}(\bar{Y}_{(\alpha-1)2}, \bar{Y}_{\alpha 2}/s) = \text{cov}\left(\frac{\Sigma Y_{(\alpha-1)i}}{Pn}, \frac{\Sigma Y_{\alpha i'}}{Qn}\right) = \text{cov}(Y_{(\alpha-1)i}, Y_{\alpha i'}/s)$$

$$= \frac{\sum_{i \neq i'} (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i'} - EY_{\alpha i'}/s)}{(n-(\alpha-1)Qn)(N-(\alpha-1)Qn-1)} = \frac{-\text{cov}(Y_{(\alpha-1)i}, Y_{\alpha i}/s)}{N-(\alpha-1)Qn-1} .$$

Therefore,

$$\underset{s}{E} \operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i}/S) = \frac{(N-(\alpha-1)Qn)!\,(\alpha-1)Qn!}{N!}$$

$$\cdot \frac{\sum_s \sum_i (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i} - EY_{\alpha i}/s)}{N-(\alpha-1)Qn}$$

Since* $\sum_s \sum_i (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i} - EY_{\alpha i}/s) =$

$$= \frac{N^2(N-2)!\,(N-(\alpha-1)Qn-1)}{(\alpha-1)Qn!\,(N-(\alpha-1)Qn)!} \quad \operatorname{cov}(Y_{(\alpha-i)i}, Y_{\alpha i}) \quad ,$$

$$\underset{s}{E} \operatorname{cov}(\overline{Y}_{(\alpha-1)2}, \overline{Y}_{\alpha 2}/s) = -\frac{N}{N-1} \; \frac{\operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i})}{N-(\alpha-1)Qn} \quad .$$

Now

$$\operatorname{cov} E(\overline{Y}_{(\alpha-1)2}/s),\ E(\overline{Y}_{\alpha 2}/s) = \operatorname{cov}\left(\frac{\Sigma\, Y_{(\alpha-1)i}}{N-(\alpha-1)Qn}\ ,\ \frac{\Sigma Y_{\alpha i'}}{N-(\alpha-1)Qn}\right)$$

$$= \frac{1}{(N-(\alpha-1)Qn)^2}\left[\sum_i \operatorname{cov}(Y_{(\alpha-i)i}, Y_{\alpha i}) + \sum_{i \neq i'} \operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i'})\right]$$

$$= \frac{1}{N-(\alpha-1)Qn}\left(1 - \frac{N-(\alpha-1)Qn-1}{N-1}\right) \operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i})$$

$$= \frac{(\alpha-1)Qn}{(N-1)(N-(\alpha-1)Qn)} \quad \operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i}) \ .$$

Hence

$$\operatorname{cov}(\overline{Y}_{(\alpha-1)2}, \overline{Y}_{\alpha 2}) = -\frac{\operatorname{cov}(Y_{(\alpha-1)i}, Y_{\alpha i})}{N-1} = \frac{-\sigma_{(\alpha-1)\alpha,1}}{N-1}$$

---

*The proof of this statement is given later.

*Proof: $\sum_s \sum_i (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i} - EY_{\alpha i}/s) =$

$$\frac{N^2(N-2)!(N-(\alpha-1)Qn-1)}{(\alpha-1)Qn!(N-(\alpha-1)Qn)!} \text{ cov}(Y_{(\alpha-1)i}, Y_{\alpha i}) .$$

Consider two random variable $Y_{(\alpha-1)i}$, $Y_{\alpha i}$, where $i \in U - S$, then

$$\text{Cov } (Y_{(\alpha-1)i}, Y_{\alpha i}) = E \text{ cov}(Y_{(\alpha-1)i}, Y_{\alpha i}/s) + \text{cov } E(Y_{(\alpha-1)i}/s), E(Y_{\alpha i}/s)$$

$$= \frac{(N-(\alpha-1)Qn)!(\alpha-1)Qn!}{N!} \frac{\sum_s \sum_i (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i} - EY_{\alpha i}/s)}{N-(\alpha-1)Qn}$$

$$+ \text{cov } \left(\frac{\Sigma Y_{(\alpha-1)i}}{N-(\alpha-1)Qn}, \frac{\Sigma Y_{\alpha i}}{N((-1)Qn}\right) .$$

Therefore

$$\sum_s \sum_i (Y_{(\alpha-1)i} - EY_{(\alpha-1)i}/s)(Y_{\alpha i} - EY_{\alpha i}/s)$$

$$= \frac{N^2(N-2)!(N-(\alpha-1)Qn-1)}{(\alpha-1)Qn!(N-(\alpha-1)Qn)!} \text{ cov } (Y_{(\alpha-1)i}, Y_{\alpha i}) .$$

Hence, we can see the covariance between two disjoint sample means on two different occasions has the same form as the covariance between two unit values on a single occasion, except that the variance of $Y_{\alpha i}$ is simply replaced by the covaraince between two unit values on successive occasions.

We also have the following relations between any two disjoint sample means on different occasions.

$$\text{cov } (\overline{Y}_{(\alpha-i)2}, \overline{Y}_{\alpha 2}) = -\frac{\sigma_{(\alpha-i)\alpha 1}}{N-1} \qquad (1 \leq i < \alpha)$$

$$\text{cov}\,(\bar{Y}_{(\alpha-i)1}, \bar{Y}_{\alpha 2}) = -\frac{\sigma_{(\alpha-i)\alpha 1}}{N-1}$$

$$\text{cov}\,(\bar{Y}_{(\alpha-i)1}, \bar{Y}_{\alpha 1}) = -\frac{\sigma_{(\alpha-i)\alpha 1}}{N-1} \quad .$$

Lemma 5.

$$\text{Cov}\,(\bar{Y}_{\alpha 1}, \bar{Y}_{\alpha 2}) = -\frac{\sigma^2_{\alpha 1}}{N-1} \quad \text{for every } \alpha \quad .$$

This is a special case of cov $(\bar{Y}_{(\alpha-i)1}, \bar{Y}_{\alpha 2})$ when $i = 0$.

In order to grasp clearly the nature of the variances and covariances between any two means on two occasions $\alpha-1$, α., summary Table 3.1 is given below. This also holds for any two arbitrary occasions.

Table 3.1. Variances and covariances between means in uni-stage sampling

	$\bar{Y}_{(\alpha-1)1}$	$\bar{Y}_{(\alpha-1)2}$	$\bar{Y}_{\alpha 1}$	$\bar{Y}_{\alpha 2}$
$\bar{Y}_{(\alpha-1)1}$	$\frac{\sigma^2_{\alpha-1,1}}{Qn}(\frac{N-Qn}{N-1})$	$-\frac{\sigma^2_{\alpha-1,1}}{N-1}$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$
$\bar{Y}_{(\alpha-1)2}$		$\frac{\sigma^2_{\alpha-1,1}}{Pn}(\frac{N-Pn}{N-1})$	$\frac{\sigma^2_{(\alpha-1)\alpha,1}}{Pn}(\frac{N-Pn}{N-1})$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$
$\bar{Y}_{\alpha 1}$			$\frac{\sigma^2_{\alpha,1}}{Pn}(\frac{N-Pn}{N-1})$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$
$\bar{Y}_{\alpha 2}$				$\frac{\sigma^2_{\alpha,1}}{Qn}(\frac{N-Qn}{N-1})$

3.3. Derivation of Variances and Covariances Multi-stage Sampling

So far, the variances and covariances between sample means on any pair of occasions have been considered for uni-stage sampling. In this section, these same quantities are derived for multi-stage sampling.

Lemma 6. Let $\hat{\bar{Y}}_{\alpha 1} = \frac{1}{Pn} \sum_i \frac{N_i}{n_i} \sum_j Y_{\alpha ij}$

$$i = (\alpha-1)Qn, \ldots, n+(\alpha-2)Qn \quad ,$$

then $\hat{\bar{Y}}_{\alpha 1}$ is an unbiased estimator of $\bar{\bar{Y}}_\alpha$

$$E\,\hat{\bar{Y}}_{\alpha 1} = \bar{\bar{Y}}_\alpha \qquad \text{for all } \alpha .$$

Lemma 7. The variance of $\hat{\bar{Y}}_{\alpha 1}$ of size Pn on occasion α is

$$V(\hat{\bar{Y}}_{\alpha 1}) = \frac{\sigma^2_{\alpha,1}}{Pn}\left(\frac{N-Pn}{N-1}\right) + \frac{\sigma^2_{\alpha,2}}{Pn} \qquad \text{for all } \alpha$$

where

$$\sigma^2_{\alpha,2} = \frac{1}{N}\sum_{i=1}^{N} N_i^2 \frac{\sigma^2_{\alpha,2i}}{n_i}\left(\frac{N_i-n_i}{N_i-1}\right)$$

Lemma 8.

$$\text{Cov}\,(\hat{\bar{Y}}_{(\alpha-1)2}, \hat{\bar{Y}}_{\alpha 1}) = \frac{\sigma_{(\alpha-1)\alpha,1}}{Pn}\left(\frac{N-Pn}{N-1}\right) + \frac{\sigma_{(\alpha-1)\alpha,2}}{Pn} \quad \text{for all } \alpha$$

where

$$\sigma_{(\alpha-1)\alpha,2} = \frac{1}{N}\sum N_i^2 \frac{\sigma_{(\alpha-1)\alpha,i}}{n_i}\left(\frac{N_i-n_i}{N_i-1}\right) \quad .$$

Lemma 9.

$$\text{cov}\,(\bar{\bar{Y}}_{(\alpha-1)1}, \bar{\bar{Y}}_{(\alpha-1)2}) = -\frac{\sigma^2_{(\alpha-1),1}}{N-1} \quad \text{for all } \alpha .$$

Proof: Let us define a subset of the preliminary sample as follows:

$$s = \{u_i: \; i = 1, 2, \ldots, (\alpha-2)Qn\}$$

$$s_1 = \{u_i: \; i = (\alpha-2)Qn+1, \ldots, (\alpha-1)Qn\}$$

$$s_2 = \{u_i: \; i = (\alpha-1)Qn+1, \ldots, n + (\alpha-2)Qn\}$$

$$s' = s + s_1 \quad ,$$

then

$$\text{cov}\,(\bar{\bar{Y}}_{(\alpha-1)1}\bar{\bar{Y}}_{(\alpha-1)2}) = E\,\text{cov}(\bar{\bar{Y}}_{(\alpha-1)1}\bar{\bar{Y}}_{(\alpha-1)2}/s') + \text{cov}(E\bar{\bar{Y}}_{(\alpha-1)1}/s', E\bar{\bar{Y}}_{(\alpha-1)2}/s') \quad .$$

First, $\operatorname{cov}\,(\bar{EY}_{(\alpha-1)1}/s'\;,\;\bar{EY}_{(\alpha-1)2}/s')$

$$= \operatorname{cov}\,\frac{\Sigma Y_{(\alpha-1)i}}{Qn}\;,\;\frac{\Sigma Y_{(\alpha-1)i}}{N-Qn} = -\frac{\sigma^2_{\alpha-1}}{N-1} \quad \text{from Lemma 5.}$$

Second, $\operatorname{cov}\,(\bar{Y}_{(\alpha-1)1}\bar{Y}_{(\alpha-1)2}/s') = \operatorname{cov}\,(\frac{1}{Qn}\,\Sigma\,\frac{N_i}{n_i}\,\Sigma Y_{(\alpha-1)ij},$

$$\frac{1}{Pn}\;\Sigma\,\frac{N_{i'}}{n_{i'}}\,\Sigma Y_{(\alpha-1)i'j}/s')$$

$$= E\,\operatorname{cov}\,(\frac{1}{Qn}\,\Sigma\,\frac{N_i}{n_i}\;\Sigma Y_{(\alpha-1)ij},\;\frac{1}{Pn}\,\Sigma\,\frac{N_{i'}}{n_{i'}}\,\Sigma\,Y_{(\alpha-1)i'j}/s',s_2)$$

$$+ \operatorname{cov}\,(E\,\frac{1}{Qn}\,\Sigma\,\frac{N_i}{n_i}\,\Sigma Y_{(\alpha-1)ij}\;,\;E\,\frac{1}{Pn}\,\Sigma\,\frac{N_{i'}}{n_{i'}}\,Y_{(\alpha-1)i'j}/s',s_2)$$

$$= 0 \quad .$$

Since, the first part is

$$\operatorname{cov}\,(Y_{(\alpha-1)ij},\;Y_{(\alpha-1)i'j}/s',s_2) = 0$$

and the second part is

$$\operatorname{cov}\,(\frac{\Sigma Y_{(\alpha-1)i}}{Qn}\;\;\frac{\Sigma Y_{(\alpha-1)i'}}{Pn}/s',s_2) = \frac{\sum_i \sum_{i'}}{PQn^2}\;\operatorname{Cov}(Y_{(\alpha-1)i}Y_{(\alpha-1)i'}/s',s_2)$$

$$= 0$$

$$\operatorname{cov}\,(\bar{Y}_{\alpha-11}\bar{Y}_{\alpha-12}) = -\frac{\sigma^2_{\alpha-1}}{N-1} \quad .$$

That is, the covariance between two sample means in multi-stage sampling is the same as that in uni-stage sampling. Similarly we can show for multi-stage sampling that

$$\operatorname{cov}\,(\bar{Y}_{\alpha-1,1}\;\bar{Y}_{\alpha 2}) = -\frac{\sigma_{(\alpha-1)\alpha\,1}}{N-1}$$

$$\text{cov}\ (\bar{Y}_{(\alpha-1)1}\ \bar{Y}_{\alpha 1}) = -\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$$

$$\text{cov}\ (\bar{Y}_{\alpha 1}\ \bar{Y}_{\alpha 2}) = -\frac{\sigma^2_{\alpha,1}}{N-1}$$

$$\text{cov}\ (\bar{Y}_{(\alpha-1)2},\ \bar{Y}_{\alpha 2}) = -\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$$

Table 3.2 gives the nature of the variances and covariances between two means on any two occasions for multi-stage sampling.

Table 3.2. Variances and covariances between means in multi-stage sampling

	$\bar{Y}_{(\alpha-1)1}$	$\bar{Y}_{(\alpha-1)2}$	$\bar{Y}_{\alpha 1}$	$\bar{Y}_{\alpha 2}$
$\bar{Y}_{(\alpha-1)1}$	$\frac{\sigma^2_{\alpha-1,1}}{Qn}\left(\frac{N-Qn}{N-1}\right)+\frac{\sigma^2_{\alpha-1,2}}{Qn}$	$-\frac{\sigma^2_{\alpha-1,1}}{N-1}$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$
$\bar{Y}_{(\alpha-1)2}$		$\frac{\sigma^2_{\alpha-1,1}}{Pn}\left(\frac{N-Pn}{N-1}\right)+\frac{\sigma^2_{\alpha-1,2}}{Pn}$	$\frac{\sigma_{(\alpha-1)\alpha,1}}{Pn}\left(\frac{N-Pn}{N-1}\right)+\frac{\sigma_{(\alpha-1)\alpha,2}}{Pn}$	$-\frac{\sigma_{(\alpha-1)\alpha,1}}{N-1}$
$\bar{Y}_{\alpha 1}$			$\frac{\sigma^2_{\alpha,1}}{Pn}\left(\frac{N-Pn}{N-1}\right)+\frac{\sigma^2_{\alpha 2}}{Pn}$	$-\frac{\sigma^2_{\alpha,1}}{N-1}$
$\bar{Y}_{\alpha 2}$				$\frac{\sigma^2_{\alpha,1}}{Qn}\left(\frac{N-Qn}{N-1}\right)+\frac{\sigma^2_{\alpha,2}}{Qn}$

4. ESTIMATION THEORY FOR SAMPLING ON TWO SUCCESSIVE OCCASIONS

4.1. Introductory Remarks

A large number of estimators of the population mean, $\bar{\bar{Y}}_\alpha$, on occasion α are available. The general estimation problem is the problem of choosing an estimator which has good properties. More precisely, our aim will be to choose an estimator for which is unbiased and has small variance. In fact, we have to restrict this thesis to several classes of estimators which seem to be efficient and practical in general sampling situations. Namely, four classes of estimators, which have been studied by others, will be considered for sampling on two occasions in this chapter and two selected classes of estimators will be discussed for sampling on more than two occasions in Chapter 5. Finally, a new sampling design will be proposed and, based on this, we will consider a new class of estimators which might be more efficient than the estimators discussed previously. This class of estimators will be examined in Chapter 6. Hence, in 4.2, the optimum estimator for each of the four classes will be determined over all possible values of Q and for $\rho > 1/2$. In 4.3, the four optimum estimators will be compared. In Section 4.4, the multi-stage sampling theory will be applied to the best estimator selected in 4.3. Finally, in Section 4.5, a symmetric estimator, which is useful for sampling on more than two occasions, will be introduced.

4.2. Estimators and Their Properties

As mentioned above, the four classes of estimates to be treated in this chapter are as follows:

class 1. General linear estimator

$${}_1\bar{Y}_2 = a\bar{Y}_{11} + b\bar{Y}_{12} + c\bar{Y}_{21} + d\bar{Y}_{22}$$

where $a + b = 0 \quad c + d = 1$

class 2. Modified composite estimator

$${}_2\bar{Y}_2 = a(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + b\bar{Y}_{22}$$

where $a + b = 1$

class 3. Composite estimator

$${}_3\bar{Y}_2 = a(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + b\bar{Y}_2$$

class 4. Ratio-type composite estimator

$${}_4\bar{Y}_2 = a(\bar{Y}_1 \frac{\bar{Y}_{21}}{\bar{Y}_{12}}) + b\bar{Y}_2$$

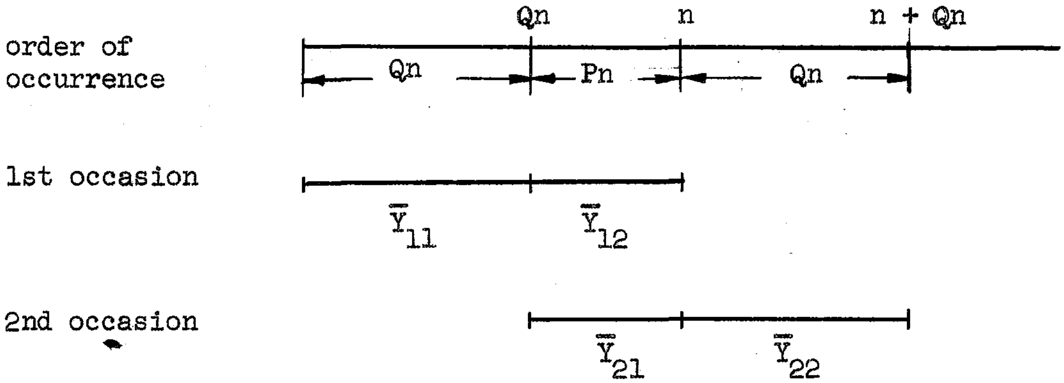
Class 1 is due to Hansen, *et al.* [5]. Class 2, which is of a form similar to the estimator given by Yates, is due to Cochran [1] when we take ${}_2\bar{Y}_1 = \bar{Y}_1$. Class 3 is due to Hansen and Hurwitz [4] when we take ${}_3\bar{Y}_1 = \bar{Y}_1$. Class 4 is due to Hansen, *et al.* [4] and Purakam [11].

The following discussion will provide derivations of the variance for each each estimator, the optimum estimator and its properties as well as the optimum values of Q for given ρ.

From Lemma 1, the estimators, ${}_1\bar{Y}_2$, ${}_2\bar{Y}_2$, and ${}_3\bar{Y}_2$ are unbiased estimators of $\bar{\bar{Y}}_2$.

4.2.1. A General Linear Estimator

First of all, let us again draw the structure of our sampling procedure for the first two occasions as follows:



The sample for the first occasion is made up of the first-stage units appearing from order 1 to n and the sample for the second occasion consists of the units appearing from $Qn + 1$ to $n + Qn$, such that the sample size n is the same on both occasions. The first stage units occurring from order 1 to Qn are in the first occasion only and the sample mean $\bar{Y}_{11}$ is obtained from those units. The first-stage units appearing from $Qn+1$ to n constitute the matched part and $\bar{Y}_{12}$ is obtained from those units on the first occasion. $\bar{Y}_{21}$ is obtained from the same units on the second occasion. These units appearing on draw n+1 to draw n+Qn constitute the part of the sample which replaced the Qn units rejected after the first occasion and $\bar{Y}_{22}$ is obtained from those units on the second occasion.

We wish to estimate $\bar{\bar{Y}}_2$ according to our sampling scheme. A general linear estimator of $\bar{\bar{Y}}_2$ is a linear function of sample means; $\bar{Y}_{11}$, $\bar{Y}_{12}$, $\bar{Y}_{21}$, $\bar{Y}_{22}$, i.e.,

$${}_1\bar{Y}_2 = a\bar{Y}_{11} + b\bar{Y}_{12} + c\bar{Y}_{21} + d\bar{Y}_{22} \quad \text{where, } a + b = 0 \quad c + d = 1 \, .$$

<u>Variance of ${}_1\bar{\bar{Y}}_2$:</u> In order to derive the variance of ${}_1\bar{\bar{Y}}_2$, we can rewrite ${}_1\bar{\bar{Y}}_2$ as

$$ {}_1\bar{\bar{Y}}_2 = a(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}) + c(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + \bar{\bar{Y}}_{22} \quad . $$

To simplify the notation, σ_{11}^2, σ_{21}^2, the variances between first-stage units on the first and second occasions respectively, will be written as σ_1^2, σ_2^2 except in Section 4.4. Then

$$ \begin{aligned} V({}_1\bar{\bar{Y}}_2) &= a^2 V(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}) + c^2(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + V(\bar{\bar{Y}}_{22}) \\ &\quad + 2ac \operatorname{cov}(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}, \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + 2a \operatorname{cov}(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}, \bar{\bar{Y}}_{22}) \\ &\quad + 2c \operatorname{cov}(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}, \bar{\bar{Y}}_{22}) \\ &= a^2 \frac{N}{N-1} \frac{\sigma^2}{nPQ} + c^2 \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} - 2ac \frac{N}{N-1} \frac{\sigma_{12}}{Pn} - 2c \frac{\sigma_2^2}{Qn} \frac{N}{N-1} \\ &\quad + \frac{\sigma_2^2}{Qn} \left(\frac{N-Qn}{N-1}\right) \quad . \end{aligned} \tag{4.1} $$

<u>The optimum estimator and its variance for given Q and ρ:</u> We are considering the optimum estimator of $\bar{\bar{Y}}_2$ among all linear estimators for given values of Q and ρ . This has a practical meaning in many cases where the replacement fraction Q is predetermined from costs or other conditions. The optimum estimator for given values of Q and ρ is the estimator having a_w, c_w given by

$$ \frac{1}{2} \frac{\partial V({}_1\bar{\bar{Y}}_2)}{\partial a} = a \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} - c \frac{N}{N-1} \frac{\sigma_{12}}{nPQ} = 0 $$

$$ \frac{1}{2} \frac{\partial V({}_1\bar{\bar{Y}}_2)}{\partial c} = -a \frac{N}{N-1} \frac{\sigma_{12}}{Pn} + c \frac{N}{N-p} \frac{\sigma_2^2}{nPQ} - \frac{N}{N-1} \frac{\sigma_2^2}{Qn} = 0 \quad . $$

This yields

$$a_w = \frac{\rho PQ}{1-Q^2\rho^2} \frac{\sigma_2}{\sigma_1}$$

$$c_w = \frac{P}{1-Q^2\rho^2}$$

Assuming that the variances on both occasions are approximately equal, the following numerical table may help to graps the nature of the weights a_w and c_w .

Table 4.1. **Optimum values of** a and c for ${}_1\overline{Y}_2$ $(\sigma_1 = \sigma_2)$

c_w = weight of $\overline{Y}_{21} - \overline{Y}_{22}$ on the second occasion (d=1-c)

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9
.6	.90	.81	.72	.64	.55	.46	.36	.26	.14
.7	.90	.81	.72	.65	.57	.49	.40	.29	.17
.8	.91	.82	.74	.67	.60	.52	.44	.34	.21
.9	.91	.83	.76	.69	.63	.57	.50	.42	.29

a_w = weight of $\overline{Y}_{11} - \overline{Y}_{12}$ on the first occasion (b = -a)

ρ \ Q									
.6	.05	.09	.13	.15	.16	.16	.15	.12	.08
.7	.06	.11	.15	.18	.20	.20	.19	.16	.10
.8	.07	.13	.18	.21	.24	.25	.25	.22	.15
.9	.08	.15	.20	.24	.28	.30	.31	.30	.24

In estimating the second occasion mean, the effect of the first occasion is quite small when ρ is less than .6 . This fact provides useful information for developing a sampling theory for more than two occasions, where it is understood that correlations between observations on the same units many occasions apart are quite small. When the sampling is based on replacement at each draw, the optimum values of a and c are the same as with the present result.

Thus, a_w and c_w can be estimated with a small bias by $\hat{a}_w$ and $\hat{c}_w$ respectively:

$$\hat{a}_w = \frac{PQr}{1-Q^2r^2} \quad \frac{s_2}{s_1}$$

$$\hat{c}_w = \frac{P}{1-Q^2r^2} \quad .$$

The proof is given in the Appendix 9.1.

For those populations in which the distribution of the observations is similar on all occasions, the difference $E\frac{s_2 - \sigma_2}{\sigma_2} - E\frac{s_1 - \sigma_1}{\sigma_1}$ will be small and the bias of $\hat{a}_w$ is approximated by

$$B = \frac{PQ\rho}{1-Q^2\rho^2} \quad \frac{\sigma_2}{\sigma_1} \; E \left(\frac{r-\rho}{\rho} + \frac{Q^2(r^2 - \rho^2)}{1-Q^2\rho^2} \right) \left(1 - \frac{(s_1 - \sigma_1)}{\sigma_1}\frac{(s_2 - \sigma_2)}{\sigma_2} + \frac{(s_1 - \sigma_1)^2}{\sigma_1^2} \right) \quad .$$

It can be seen that the bias is maximum when $E\frac{(s_1 - p_1)(s_2 - \sigma_2)}{\sigma_1\sigma_2} = 0$ and the bias is reduced when $E\frac{(s_1 - \sigma_1)(s_2 - \sigma_2)}{\sigma_1\sigma_2} = E\frac{(s_1 - \sigma_1)^2}{\sigma_1^2}$.

The bias of $\hat{a}_w$ is shown below for some values of ρ and n .

Table 4.2. Maximum bias of $\hat{a}_w$ where $P = .5$, $\sigma_1 = \sigma_2$

ρ \ n	10	30	100	a_w
.7	.0159	.0054	.0016	.20
.8	.0201	.0067	.0020	.24
.9	.0256	.0085	.0025	.28

Table 4.3. Bias of $\hat{a}_w$ when $P = .5$, $\sigma_1 = \sigma_2$, $E\dfrac{(s_1 - \sigma_1)(s_2 - \sigma_2)}{\sigma_1\sigma_2} = E\dfrac{(s_1-\sigma_1)^2}{\sigma_1^2}$

ρ \ n	10	30	100	a_w
.7	.0035	.0012	.0004	.20
.8	.0033	.0011	.0003	.24
.9	.0023	.0008	.0002	.28

On the average, the relative bias of $\hat{a}_w$ to a_w is less than 5% when $n = 10$. The relative bias of $\hat{c}_w$ to c_w is less than 1% when $n = 10$.

Table 4.4. Bias of $\hat{c}_w$ when $p = .5$

ρ \ n	10	30	100	a_w
.7	.0042	.0014	.0004	.57
.8	.0022	.0007	.0002	.60
.9	.0007	.0002	.0001	.63

The optimum estimator of $\bar{\bar{Y}}_2$ (within the general linear class) is

$${}_1\bar{Y}_w = a_w\bar{Y}_{11} + (1-a_w)\bar{Y}_{12} + c_w\bar{Y}_{21} + (1-c_w)\bar{Y}_{22}$$

and its variance obtained by putting a_w, c_w in (4.1) is

$$V({}_1\bar{Y}_w) = \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ}\,\frac{\rho^2}{1-Q^2\rho^2} \quad . \tag{4.2}$$

Other alternative forms are

$$= \frac{\sigma_2^2}{n}\left(\frac{1-Q\rho^2}{1-Q^2\rho^2}\right)\frac{N}{N-1} - \frac{\sigma_2^2}{N-1} \tag{4.3}$$

$$= \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right)\left(\frac{1-Q\rho^2}{1-Q^2\rho^2}\right) - a_w\,\frac{\sigma_{12}}{N-1} \tag{4.4}$$

$$= \frac{\sigma_2^2}{n}\left(\frac{1-Q\rho^2}{1-Q^2\rho^2}\right) - \frac{\sigma_2^2}{N-1}\left(1 - \frac{1-Q\rho^2}{n(1-Q^2\rho^2)}\right) \quad . \tag{4.5}$$

Each formula has its own purpose; namely, (4.2) serves for comparison among the four estimates which will be given in Section 4.3. Equation (4.4) gives the comparison between this and the estimator of $\bar{\bar{Y}}_2$ from a sample on the second occasion only. As we can see, even neglecting the reduction factor $\frac{a_w\sigma_{12}}{N-1}$ in (4.4), the optimum estimator on successive occasions is still smaller than the variance of $\bar{\bar{Y}}_2$ on a single occasion. Equation (4.5) gives a basis for comparison with sampling with replacement at each draw. The optimum variance of ${}_1\bar{\bar{Y}}_2$ with replacement at each draw is given as

$$V\,({}_1\bar{Y}_2)_{opt} = \frac{\sigma_2^2}{n}\left(\frac{1-Q\rho^2}{1-Q^2\rho^2}\right) \quad .$$

Therefore, the further reduction in the optimum variance based on sampling without replacement of units at each draw is

$$\frac{\sigma_2^2}{N-1}\left(1 - \frac{1-Q\rho^2}{n(1-Q^2\rho^2)}\right)$$

$V({}_1\bar{Y}_w)$ can be estimated by $\hat{V}({}_1\bar{Y}_w)$

$$\hat{V}\ ({}_1\bar{Y}_w) = \frac{s_2^2}{n}\ (\frac{1-Qr^2}{1-Q^2r^2}) - \frac{s_2^2}{N} \quad .$$

The proof is given in 9.1.3.

From 9.1.3, the relative bias of $\hat{V}({}_1\bar{Y}_w)$ to $V({}_1\bar{Y}_w)$ when $Q = .5$ for selected values of ρ and n is given below.

Table 4.5. Relative bias of $\hat{V}({}_1\bar{Y}_w)$ to $V({}_1\bar{Y}_w)$ when $Q = .5$

ρ \ n	10	30	100
.7	.0098	.0032	.0009
.8	.0062	.0021	.0005
.9	.00019	.0006	.0001

In practical applications, the estimate of ${}_1\bar{Y}_w$, ${}_1\tilde{\bar{Y}}_w$, is considered:

$${}_1\tilde{\bar{Y}}_w = \hat{a}_w\ \bar{Y}_{11} + (1-\hat{a}_w)\bar{Y}_{12} + \hat{c}_w\bar{Y}_{21} + (1-\hat{c}_w)\ \bar{Y}_{22} \quad .$$

The variance of ${}_1\tilde{\bar{Y}}_w$ is approximately the same as the variance of ${}_1\bar{Y}_w$ as shown in the Appendix 9.1.4.

<u>Optimum Q for given ρ</u>: The optimum value of Q for given ρ can be obtained by

$$-\frac{n(N-1)(1-Q^2\rho^2)^2}{N\ \sigma^2}\ \frac{d}{dQ}\ V({}_1\bar{Y}_w) = Q^2\rho^2 - 2Q+1 = 0 \quad .$$

Hence, $Q_{opt} = \dfrac{1-\sqrt{1-\rho^2}}{\rho^2}$. This is the same result as for sampling with replacement at each draw. Consider the limiting value of the optimum Q.

When $\rho = 1$, the optimum Q is equal to 1. However, when ρ approaches zero,

$$\lim_{\rho \to 0} Q_{opt} = \lim_{\rho \to 0} \frac{1}{2} \cdot \frac{1}{\sqrt{1-\rho^2}} = \frac{1}{2} \quad .$$

Therefore, the optimum Q is always larger than .5 for all values of ρ.

The variance of ${}_1\bar{Y}_2$ at the optimum values of Q, a, and c is :

$$V({}_1\bar{Y}_w) = \frac{\sigma_2^2}{n} \left(\frac{N-n}{N-1}\right) \frac{\rho^2}{2(1-\sqrt{1-\rho^2})} - \frac{\sigma_2^2}{N-1} \left(1 - \frac{\rho^2}{2(1-\sqrt{1-\rho^2})}\right)$$

$$= \frac{\sigma_2^2}{n} \quad \frac{\rho^2}{2(1-\sqrt{1-\rho^2})} - \frac{\sigma_2^2}{N-1} \left(1 - \frac{\rho^2}{2n(1-\sqrt{1-\rho^2})}\right) \quad . \quad (4.6)$$

This can be compared with the optimum variance for sampling with replacement at each draw, which is

$$V({}_1\bar{Y}_2)_{opt} = \frac{\sigma_2^2}{n} \quad \frac{\rho^2}{2(1-\sqrt{1-\rho^2})} \quad .$$

In order to have some idea of the optimum Q and the efficiency of the estimator relative to the sample mean on the single occasion ($\bar{Y}_2$), Table 4.6 has been prepared. Here, for simplicity, the reduction terms in (4.5) and (4.6) are neglected. The percent gain in efficiency of ${}_1\bar{Y}_w$ over $\bar{Y}_2$ is defined as

$$\frac{\text{Var}(\bar{Y}_2) - \text{Var}({}_1\bar{Y}_w)}{\text{Var}({}_1\bar{Y}_w)} \cdot 100 \quad .$$

The important result is that the optimum value of Q (discarded fraction) is always larger than .5 for all values of ρ, and it becomes larger when the correlation between observations on the same unit is higher. This implies

Table 4.6. The percent gain of ${}_1\bar{Y}_w$ over $\bar{Y}_2$ and optimum Q

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9	Opt Q	Gain at opt Q
.6	3.36	6.20	8.47	10.09	10.97	11.02	10.10	8.08	4.79	.55	11.11
.7	4.68	8.69	12.06	14.62	16.22	16.65	15.66	12.89	7.88	.58	16.73
.8	6.15	11.74	16.63	20.64	23.50	24.93	24.34	20.98	13.58	.62	25.00
.9	7.93	15.46	22.47	28.75	34.03	37.82	39.28	36.81	26.90	.69	39.50
.95	8.92	17.63	25.98	33.89	41.11	47.24	51.45	51.94	43.24	.77	53.33

that in order to estimate $\bar{\bar{Y}}_2$ on successive occasions the higher the correlation, the smaller the portion matched. In fact, when the correlation between the two occasions is high, the matched portion on the second occasion provides approximately the same amount of information as it did on the first occasion. Therefore, we can gain more information overall with respect to the population values on the second occasion by taking more new units into the sample on the second occasion when ρ is high.

4.2.2. A Modified Composite Estimator

A modified composite estimator is a composite of two estimators. The first is the sample mean based on the Qn units selected on the second occasion only. The second is the sample mean for the first occasion, to which has been added an estimate of the change from the first occasion to the second occasion. The form of the modified composite estimator is

$$ {}_2\bar{Y}_2 = a(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + b\bar{Y}_{22} $$

$$ \text{where} \quad a + b = 1 . $$

As before, its variance and optimum properties will be considered.

<u>Variance of ${}_2\bar{\bar{Y}}_2$:</u> ${}_2\bar{\bar{Y}}_2$ can be written as

$${}_2\bar{\bar{Y}}_2 = aQ(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}) + a(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + \bar{\bar{Y}}_{22} \quad .$$

From Chapter 3 we have

$$V({}_2\bar{\bar{Y}}_2) = a^2Q^2V(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}) + a^2\ V(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + V(\bar{\bar{Y}}_{22})$$

$$+ 2\ a^2Q\ \text{cov}\ (\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}, \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + 2\ aQ\ \text{cov}(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}, \bar{\bar{Y}}_{22})$$

$$+ 2a\ \text{cov}(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}, \bar{\bar{Y}}_{22})$$

$$= a^2Q^2\ \frac{N}{N-1}\ \frac{\sigma_1^2}{nPQ} + a^2\ \frac{N}{N-1}\ \frac{\sigma_2^2}{nPQ} + \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right)$$

$$- 2\ a^2Q\ \frac{N}{N-1}\ \frac{\sigma_{12}}{Pn} - 2\ a\ \frac{N}{N-1}\ \frac{\sigma_2^2}{Qn}$$

$$= a^2\ \frac{N}{N-1}\ \frac{1}{nPQ}\ [Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}] - 2a\ \frac{N}{N-1}\ \frac{\sigma_2^2}{Qn} + \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) \tag{4.7}$$

<u>Optimum estimator and its variance for given Q and ρ:</u>

$$\frac{(N-1)nPQ}{2N}\ \frac{d}{da}\ V({}_2\bar{\bar{Y}}_2) = a(Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}) - P\sigma_2^2 = 0 \quad .$$

Hence,

$$a_w = \frac{P\sigma_2^2}{Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}}$$

under $\sigma_1 = \sigma_2$,

$$a_w = \frac{P}{Q^2 + 1 - 2Q^2\rho} \quad .$$

Table 4.7. Optimum values of $\underline{a}$ for ${}_2\bar{Y}_2$ $(\sigma_1 = \sigma_2)$

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9
.6	.90	.81	.71	.62	.53	.43	.33	.23	.12
.7	.90	.81	.73	.64	.56	.47	.37	.27	.15
.8	.91	.82	.74	.66	.59	.51	.42	.32	.19
.9	.91	.83	.75	.69	.63	.56	.49	.41	.28
.95	.91	.83	.76	.70	.65	.59	.54	.47	.37

The optimum values of $\underline{a}$ depend mainly on the value of Q and as the matched portion P becomes large, a_w becomes large. The a_w has a similar pattern to that of c_w of ${}_1\bar{Y}_1$ in Table 4.1.

The optimum estimator of ${}_2\bar{Y}_2$ for given Q and ρ is

$${}_2\bar{Y}_w = a_w(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + (1-a_w)\bar{Y}_{22}$$

and its variance is

$$V({}_2\bar{Y}_w) = \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ}\left(\frac{P^2\sigma_2^2}{Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}}\right) \tag{4.8}$$

$$\text{or} \quad = \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right)\left(\frac{Q\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}}{Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}}\right) - \frac{\sigma_2^2}{N-1} \cdot \frac{PQ(2\sigma_{12} - \sigma_1^2)}{Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}} \tag{4.9}$$

Formula (4.8) will be used for the comparisons among our four classes of estimates. Now, assuming $\sigma_1 = \sigma_2$ in (4.9) the optimum variance is smaller than the variance of $\bar{Y}_2$ if $\rho > \frac{1}{2}$. This follows because

$$\frac{Q + 1 - 2Q\rho}{Q^2 + 1 - 2Q^2\rho} < 1$$

if and only if

$$Q + 1 - 2Q\rho < Q^2 + 1 - 2Q^2\rho \quad ,$$

that is,

$$Q(1-2\rho) < Q^2(1 - 2\rho) \quad .$$

But this holds whenever $1 - 2\rho < 0$ and in this case we also have

$$\frac{2\rho - 1}{Q^2 + 1 - 2Q^2\rho} > 0 \quad \text{for all} \quad Q \quad .$$

Otherwise $V({}_2\bar{Y}_w) > V(\bar{Y}_2)$.

<u>Optimum Q for given ρ</u>: In order to determine the effect of an approximate optimum Q, assume $\sigma_1 = \sigma_2$. Then (4.8) can be written as

$$V({}_2\bar{Y}_w) = \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} \left(P - \frac{P^2\sigma_2^2}{Q^2\sigma_1^2 + \sigma_2^2 - 2Q^2\sigma_{12}}\right) - \frac{\sigma_2^2}{N-1}$$

$$\doteq \frac{N}{N-1} \frac{\sigma_2^2}{n} \left(\frac{Q + 1 - 2Q\rho}{Q^2 + 1 - 2Q^2\rho}\right) - \frac{\sigma_2^2}{N-1} \quad . \qquad (4.10)$$

Hence, the optimum Q for given ρ can be obtained as follows:

$$\frac{(N-1)n}{N\ \sigma_2^2} \frac{d}{dQ} V({}_2\bar{Y}_w) = \frac{(1-2\rho)(Q^2 + 1 - 2Q^2\rho) - (Q+1-2Q\rho)(2Q-4Q\rho)}{(Q^2 + 1 - 2Q^2\rho)^2} = 0$$

which yields

$$Q^2 + 1 - 2Q^2\rho - 2Q^2\rho - 2\rho + 4Q^2\rho^2 - 2Q^2$$

$$- 2Q + 4Q^2\rho - 2Q^2\rho + 4Q\rho - 8Q^2\rho^2 = 0$$

implying

$$Q^2(2\rho-1)^2 - 2Q(2\rho-1) + (2\rho-1) = 0$$

$$Q^2(2\rho-1) - 2Q+1 = 0 \quad .$$

Therefore,

$$Q_{opt} = \frac{1 - \sqrt{2 - 2\rho}}{2\rho - 1} \quad .$$

Consider the limiting value of Q_{opt} in the following cases: when $\rho = 1$, optimum Q is equal to 1. However, when ρ approaches .5

$$\lim_{\rho \to .5} Q_{opt} = \lim_{\rho \to .5} \frac{1}{2\sqrt{2-2\rho}} = \frac{1}{2} \quad .$$

When $\rho = 0$, the optimum Q is equal to .414 which is less than .5. Hence, the optimum Q lies between .4 and 1, and is always larger than .5 when $\rho > .5$.

Table 4.8 gives the percent gain of ${}_2\overline{Y}_w$ over $\overline{Y}_2$ and the optimum Q assuming $\sigma_1 = \sigma_2$ and neglecting the reduction term in (4.9).

Table 4.8. The percent gain of ${}_2\overline{Y}_w$ over $\overline{Y}_2$ and optimum Q ($\sigma_1 = \sigma_2$)

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9	Opt Q	Gain at opt Q
.6	1.83	3.33	4.46	5.21	5.55	5.45	4.88	3.80	2.19	.53	5.57
.7	3.75	6.95	9.54	11.42	12.50	12.63	11.66	9.41	5.62	.56	12.69
.8	5.74	10.90	15.36	18.94	21.42	22.50	21.72	18.46	11.73	.61	22.50
.9	7.83	15.23	22.10	28.23	33.33	36.92	38.18	35.55	25.61	.69	38.19
.95	8.90	17.56	25.89	33.75	40.90	46.95	51.08	51.42	42.63	.76	51.69

The percent gain of ${}_2\overline{Y}_w$ over $\overline{Y}_2$ is defined as before

$$\frac{V(\overline{Y}_2) - V({}_2\overline{Y}_w)}{V({}_2\overline{Y}_w)} \cdot 100 = \frac{PQ(2\rho-1)}{1-Q(2\rho-1)} \quad .$$

Here again we see that the optimum Q for the range of ρ considered, is larger than .5 and is directly related to the correlation coefficient.

4.2.3. A Composite Estimator $_3\bar{Y}_2$

A composite estimator $_3\bar{Y}_2$ is of the same form as the modified composite estimator, but uses the sample mean on the second occasion $\bar{Y}_2$ instead of $\bar{Y}_{22}$. Notationally,

$$_e\bar{Y}_2 = a(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + b\,\bar{Y}_2$$

$$\text{where} \quad a + b = 1 \quad .$$

Variance of $_3\bar{Y}_2$: $_3\bar{Y}_2$ can be written as

$$_3\bar{Y}_2 = a(Q\bar{Y}_{11} + P\bar{Y}_{12} + \bar{Y}_{21} - \bar{Y}_{12}) + (1-a)(P\bar{Y}_{21} + Q\bar{Y}_{22})$$

$$= aQ\bar{Y}_{11} - aQ\bar{Y}_{12} + (1-Q + aQ)\bar{Y}_{21} + (Q-aQ)\bar{Y}_{22}$$

$$= \bar{Y}_{21} + aQ(\bar{Y}_{11} - \bar{Y}_{12}) + (Q-aQ)(\bar{Y}_{22} - \bar{Y}_{21})$$

$$V(_3\bar{Y}_2) = \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) + a^2Q^2\,\frac{N}{N-1}\,\frac{\sigma_1^2}{nPQ} + (Q-aQ)^2\,\frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ}$$

$$- 2aQ\,\frac{N}{N-1}\,\frac{\sigma_{12}}{Pn} - 2(Q-aQ)\,\frac{N}{N-1}\,\frac{\sigma_2^2}{Pn} + 2aQ(Q-aQ)\,\frac{N}{N-1}\,\frac{\sigma_{12}}{Pn}$$

$$= \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) + a^2Q^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\,(\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12})$$

$$- 2aQ^2\,\frac{N}{N-1}\,\frac{P}{nPQ}\,\sigma_{12} - Q^2\,\frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ} \qquad (4.11)$$

$$= \frac{\sigma_2^2}{n}\left(\frac{N-n}{n-1}\right) + a^2Q^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\,(\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}) - 2aQ^2P\,\frac{N}{N-1}\,\frac{\sigma_{12}}{nPQ} \quad .$$

The variances and covariances needed in $V(_3\bar{Y}_2)$ are as follows.

$$V(\bar{Y}_{11} - \bar{Y}_{12}) = \frac{N}{N-1}\,\frac{\sigma_1^2}{nPQ}$$

$$V(\bar{Y}_{22} - \bar{Y}_{21}) = \frac{N}{N-1} \frac{\sigma_2^2}{nPQ}$$

$$Cov(\bar{Y}_{11} - \bar{Y}_{12}, \bar{Y}_{22} - \bar{Y}_{21}) = \frac{N}{N-1} \frac{\sigma_{12}}{Pn}$$

$$Cov(\bar{Y}_{21}, \bar{Y}_{11} - \bar{Y}_{12}) = \frac{-N}{N-1} \frac{\sigma_{12}}{Pn}$$

$$Cov(\bar{Y}_{21}, \bar{Y}_{22} - \bar{Y}_{21}) = -\frac{N}{N-1} \frac{\sigma_2^2}{Pn} \quad .$$

Optimum estimator and its variance for given Q and ρ: The optimum value of a is obtained as

$$a_w = \frac{\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}}$$

under the assumption $\sigma_1 = \sigma_2$, $a_w = \frac{P\rho}{2(1-Q\rho)}$

Table 4.9. Optimum values of a for $_3\bar{Y}_2$ ($\sigma_1 = \sigma_2$)

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9
.6	.29	.27	.26	.24	.21	.19	.16	.12	.07
.7	.34	.33	.31	.29	.27	.24	.21	.16	.07
.8	.39	.38	.37	.35	.33	.31	.27	.22	.14
.9	.45	.44	.43	.42	.41	.39	.36	.32	.24
1.0	.50	.50	.50	.50	.50	.50	.50	.5	.50

In contrast with the optimum values of a for $_2\bar{Y}_w$ in Table 4.7, a_w for $_3\bar{Y}_2$ is always less than .5. This implies that the sample mean on the second occasion, $\bar{Y}_2$, dominates the estimator $_3\bar{Y}_2$.

The optimum estimator of $\bar{\bar{Y}}_2$ in class 3 is

$${}_3\bar{\bar{Y}}_w = a_w(\bar{\bar{Y}}_1 + \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{12}) + (1-a_w)\bar{\bar{Y}}_2$$

and the variance of ${}_3\bar{\bar{Y}}_w$ for given Q and ρ is

$$V({}_3\bar{Y}_w) = \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\;\frac{Q^2}{nPQ}\;\frac{\sigma_2^2(\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12} + P^2\rho^2\sigma_1^2)}{\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}} \quad . \tag{4.12}$$

Since $N - Pn = (N-n) + nQ$, $N = (N-n) + n$, we can rewrite $V({}_3\bar{Y}_w)$ as:

$$V({}_3\bar{Y}_w) = \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right)\frac{\sigma_1^2(1-PQ\rho^2) + \sigma_2^2 - 2Q\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}} - \frac{\sigma_{12}^2}{N-1}\;\frac{PQ}{\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}} \tag{4.13}$$

$$= \frac{N}{N-1}\;\frac{\sigma_2^2}{n}\left(\frac{\sigma_1^2(1-PQ\rho^2) + \sigma_2^2 - 2Q\sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2Q\sigma_{12}}\right) - \frac{\sigma_2^2}{N-1} \quad . \tag{4.14}$$

From (4.13), we see that the variance of ${}_3\bar{Y}_w$ is smaller than the variance of $\bar{Y}_2$, because $(1 - PQ\rho^2) < 1$.

<u>Optimum Q for given ρ</u>: From (4.14), the optimum Q is obtained as

$$Q_{opt} = \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 2\sigma_{12}(\sigma_1^2 + \sigma_2^2}}{2\sigma_{12}} \quad .$$

In order to determine an approximate optimum Q, assume $\sigma_1 = \sigma_2$

$$Q = \frac{2 \pm \sqrt{4 - 4\rho}}{2\rho} = \frac{1 - \sqrt{1-\rho}}{\rho}$$

$$\therefore\; Q_{opt} \doteq \frac{1 - \sqrt{1-\rho}}{\rho} \quad .$$

Consider the limiting values of the optimum Q. When $\rho = 1$, optimum Q is equal to 1, and when ρ approaches zero,

$$\lim_{\rho \to 0} Q_{opt} = \lim_{\rho \to 0} \frac{1}{2\sqrt{1-\rho}} = \frac{1}{2} \quad .$$

Hence again the optimum Q is always larger than .5 . Table 4.10 gives the percent gain of ${}_3\overline{\overline{Y}}_w$ over $\overline{\overline{Y}}_2$ and optimum Q values, assuming $\sigma_1 = \sigma_2$ and neglecting the reduction term in (4.13).

Table 4.10. Percent gain of ${}_3\overline{\overline{Y}}_w$ over $\overline{\overline{Y}}_2$ and optimum Q ($\sigma_1 = \sigma_2$)

ρ \ Q	.1	.2	.3	.4	.5	.6	.7	.8	.9	Opt Q	Gain at opt Q
.6	1.75	3.38	4.83	6.02	6.87	7.23	6.97	5.86	3.65	.61	7.23
.7	2.42	4.77	6.96	8.89	10.40	11.28	11.22	9.78	6.33	.64	11.38
.8	3.23	6.49	9.69	12.73	15.38	17.32	18.02	16.58	11.46	.69	18.01
.9	4.17	8.58	13.18	17.90	22.56	26.79	29.84	30.11	23.73	.76	30.50
.95	4.69	9.78	15.27	21.16	27.36	33.64	39.43	43.02	38.88	.82	43.59

The percentage gain ${}_3\overline{\overline{Y}}_w$ over $\overline{\overline{Y}}_2$ is

$$G(\%) = \frac{V(\overline{\overline{Y}}_2) - V({}_3\overline{\overline{Y}}_w)}{V({}_3\overline{\overline{Y}}_w)} \doteq \frac{PQ\rho^2}{2(1-Q\rho)-PQ\rho^2} \quad .$$

The optimum values of Q are always greater than .5 and directly related to the correlation coefficient.

4.2.4. A Composite Ratio Estimator ${}_4\overline{\overline{Y}}_2$

The composite ratio estimator is also a composite of two estimators. The first is the sample mean based on the n units selected on the second occasion only, and the second is the sample mean for the first occasion multiplied by the ratio of the means for the first occasion to the second occasion for the matched units only. This estimator is a biased estimator of $\bar{\bar{Y}}_2$, and somewhat different from the previous three unbiased estimators,

$$_4\bar{Y}_2 = a \left(\bar{Y}_1 \cdot \frac{\bar{Y}_{21}}{\bar{Y}_{12}}\right) + b\, \bar{Y}_2$$

$$\text{where} \quad a + b = 1 \quad .$$

The bias is given in the Appendix 9.2.

<u>The variance of $_4\bar{Y}_2$</u> is quite complicated. An approximation of the variance of $_4\bar{Y}_2$ seems to be useful in practice. Thus,

$$V(_4\bar{Y}_2) \doteq \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right) + a^2Q^2\, \frac{N}{N-1}\, \frac{1}{nPQ}\, [R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}]$$

$$- 2aR\, \frac{N}{N-1}\, \frac{1}{nPQ}\, PQ^2\sigma_{12} \tag{4.15}$$

$$\text{where} \quad R = \frac{\bar{\bar{Y}}_2}{\bar{\bar{Y}}_1} \quad .$$

The derivation is given in the Appendix 9.3.

Notice that $V(_4\bar{Y}_2) \doteq V(_3\bar{Y}_2))$ if $R = 1$. That is, if the population means for both occasion are the same, then variance of the composite estimator is approximately the same as that of the composite ratio estimator.

<u>Optimum estimator and its variance for given Q and ρ:</u> As before, the optimum value of <u>a</u> is obtained from

$$\frac{(N-1)nPQ}{2NQ^2}\, \frac{d}{da}\, V(_4\bar{Y}_2) = a\, [R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}] - RP\sigma_{12} = 0$$

$$a_w = \frac{PR\sigma_{12}}{R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}} \quad .$$

Again, the optimum value of $\underline{a}$ is the same for both estimator ${}_3\bar{\bar{Y}}_2$, ${}_4\bar{\bar{Y}}_2$ if $R = 1$. The optimum estimator of ${}_4\bar{\bar{Y}}_2$ is

$$ {}_4\bar{\bar{Y}}_w = a_w(\bar{\bar{Y}}_1 \frac{\bar{\bar{Y}}_{21}}{\bar{\bar{Y}}_{12}}) + (1-a_w)\, \bar{\bar{Y}}_2 $$

and its variance is

$$ V({}_4\bar{\bar{Y}}_w) \doteq \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right) - \frac{R^2PQ}{R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}} \frac{N}{N-1} \frac{\sigma_{12}^2}{n} \qquad (4.16) $$

or

$$ \doteq \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right)\left(\frac{(1-\rho^2PQ)R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}}{R^2\sigma_1^2 + \sigma_2^2 - 2RQ\sigma_{12}}\right) - a_w \frac{RQ\sigma_{12}}{N-1} \qquad (4.17) $$

Since $1 - \rho^2 PQ < 1$, we have

$$ V({}_4\bar{\bar{Y}}_w) < V(\bar{\bar{Y}}_2) \quad \text{for all values of } P,\ \rho\ . $$

<u>Optimum Q for given ρ</u>: Under the assumption that $\sigma_1 = \sigma_2$, to simplify the computation, the optimum value of Q is given by

$$ (R^2 - 2R^2Q)(R^2+1 - 2RQ\rho)\sigma_1^2 + (R^2Q - R^2Q^2)2R\rho\sigma_1^2 = 0 $$

implying $\quad (1 - 2Q)(R^2 + 1 - 2RQ\rho) + Q(1-Q)2R\rho = 0$

which is

$$ Q^2 2R\rho - 2Q(1+R^2) + (1+R^2) = 0 $$

$$ Q = \frac{2(1+R^2) \pm \sqrt{4(1+R^2)^2 - 8R\rho(1+R^2)}}{4R\rho} . $$

Therefore, the optimum Q is

$$ Q_{opt} = \frac{(1+R^2) - \sqrt{(1+R^2)^2 - 2R\epsilon(1+R^2)}}{2R\rho} $$

If $R = 1$, then

$$Q_{opt} = \frac{1 - \sqrt{1-\rho}}{\rho}$$

which is the same as for the estimator ${}_3\bar{Y}_w$. So far, we have investigated approximate properties or the composite ratio estimator. Moreover, we have found that both the composite estimator and the composite ratio estimator have the same properties when $\bar{\bar{Y}}_1 = \bar{\bar{Y}}_2$ with respect to the variance, the optimum weight $\underline{a}$ and the optimum discarded fraction Q .

4.3. Comparison of Estimators

In 4.2, we have investigated the properties of four classes of estimators. Now, it is necessary to compare the efficiency among these four estimators. In fact, it was already found that a general linear estimator is the best among the four estimators compared to $\bar{Y}_2$.

In this section, if we express the four estimates in the following form, then we show that the general linear estimator is indeed a general form for the four estimators, and is the best among our four classes of estimators. Let

$$ {}_1\bar{Y}_2 = a_1\bar{Y}_{11} + b_1\bar{Y}_{12} + c_1\bar{Y}_{21} + d_1\bar{Y}_{22} \quad \text{where } a_1 + b_1 = 0 $$

$$ c_1 + d_1 = 1 $$

then

$$ {}_2\bar{Y}_2 = a(\bar{Y}_1 + \bar{Y}_{21} - \bar{Y}_{12}) + b\bar{Y}_{22} \qquad a + b = 1 $$

$$ = a(P\bar{Y}_{12} + Q\bar{Y}_{11} + \bar{Y}_{21} - \bar{Y}_{12}) + (1-a)\bar{Y}_{22} $$

$$ = aQ\bar{Y}_{11} - aQ\bar{Y}_{12} + a\bar{Y}_{21} + (1-a)\bar{Y}_{22} $$

$$ = a_2\bar{Y}_{11} + b_2\bar{Y}_{12} + c_2\bar{Y}_{21} + d_2\bar{Y}_{22} $$

where $a_2 + b_2 = 0$, $c_2 + d_2 = 1$,

$a_2 + Qc_2$.

This estimator has an additional restriction over the general linear estimator. Hence, the class 2 type of estimators is a subclass of the class 1 type of estimators.

$$_3\bar{\bar{Y}}_2 = a\ (\bar{\bar{Y}}_1 + \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{12}) + b\bar{\bar{Y}}_2 \qquad a + b = 1$$

$$= a(P\bar{\bar{Y}}_{12} + Q\bar{\bar{Y}}_{11} + \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{12}) + (1-a)(P\bar{\bar{Y}}_{21} + Q\bar{\bar{Y}}_{22})$$

$$= aQ\bar{\bar{Y}}_{11} - aQ\bar{\bar{Y}}_{12} + (1-bQ)\bar{\bar{Y}}_{21} + bQ\bar{\bar{Y}}_{22}$$

$$= a_3\bar{\bar{Y}}_{11} + b_3\bar{\bar{Y}}_{12} + c_3 Y_{21} + a_3\bar{\bar{Y}}_{22}$$

where $a_3 = aQ$ $\quad c_3 = 1 - bQ$,

where $a_3 + b_3 = 0$, $c_3 + d_3 = 1$

$d_3 - b_3 = Q$.

This estimator has an additional restriction over the general linear estimator. Hence the class 2 type of estimators is a subclass of the class 1 type of estimates. A composite ratio estimator is the only estimator which does not belong to the class of linear unbiased estimators. But under some restrictions on the observations, we can reduce the composite ratio estimator to a linear estimator which is a special case of the general linear estimator.

$$_4\bar{\bar{Y}}_2 = a\ (\bar{\bar{Y}}_1 \cdot \frac{\bar{\bar{Y}}_{21}}{\bar{\bar{Y}}_{22}}) + b\bar{\bar{Y}}_2$$

$$= a(P\bar{\bar{Y}}_{12} + Q\bar{\bar{Y}}_{11}) \frac{\bar{\bar{Y}}_{21}}{\bar{\bar{Y}}_{12}} + b\ (P\bar{\bar{Y}}_{21} + Q\bar{\bar{Y}}_{22})$$

$$= a(P\bar{Y}_{21} + Q\,\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}) + bP\bar{Y}_{21} + bQ\bar{Y}_{22}$$

$$= aQ\,(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}) + P\bar{Y}_{21} + bQ\bar{Y}_{22} \quad .$$

From Appendix 9.2

$$\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} = \bar{\bar{Y}}_2(1 + \Delta\bar{Y}_{11} + \Delta\bar{Y}_{21} + \Delta\bar{Y}_{11}\Delta Y_{21})(1 - \Delta\bar{Y}_{12})$$

$$+ \frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} \quad \frac{(\bar{Y}_{12} - \bar{\bar{Y}}_1)^2}{\bar{\bar{Y}}_1^2} \quad .$$

Ignoring the cross product terms

$$\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} \doteq \bar{Y}_2\ (1 + \Delta\bar{Y}_{11} + \Delta\bar{Y}_{21} - \Delta\bar{Y}_{12}) \quad .$$

Hence

$${}_4\bar{Y}_2 = aQ\,(\bar{\bar{Y}}_2 + R(\bar{Y}_{11} - \bar{\bar{Y}}_1) + (\bar{Y}_{21} - \bar{\bar{Y}}_2) - R(\bar{Y}_{12} - \bar{\bar{Y}}_1) + P\bar{Y}_1) + P\bar{Y}_{21} + bQ\bar{Y}_{22}$$

$$= aQR\bar{Y}_{11} - aQR\bar{Y}_{12} + (1 - bQ)\bar{Y}_{21} + bA\bar{Y}_{22}$$

$$= a_4\bar{Y}_{11} + b_4\bar{Y}_{12} + c_4\bar{Y}_{21} + d_4\bar{Y}_{22} \quad ,$$

where

$$a_4 + b_4 = 0 \ , \quad c_4 + d_4 = 1$$

$$d_4R - b_4 = QR \quad .$$

This estimator has an additional restriction over the general linear estimators. Hence the class 2 type of estimators is a subclass of the class 1 type of estimators. Also, if R=1, ${}_4\bar{Y}_2$ is the same as ${}_3\bar{Y}_2$.

In conclusion, the class of ${}_1\bar{Y}_2$ is a wider class of estimators among the classes under consideration. Hence the optimum estimator in this wider class should be uniformly better than any of the others.

4.4. A General Linear Estimator in Multi-stage Sampling

In the previous sections, we observed that a general linear estimator is the best among the four classes of estimators for estimating the population mean on the second occasion when uni-stage sampling is used. Therefore, a theory of estimation for multi-stage sampling on two occasions will be developed only for the general linear estimator class.

A general linear estimator ${}_1\hat{\hat{Y}}_2$ for multi-stage sampling is defined as

$$ {}_1\hat{\hat{Y}}_2 = a\hat{\hat{Y}}_{11} + b\hat{\hat{Y}}_{12} + c\hat{\hat{Y}}_{21} + d\hat{\hat{Y}}_{22} $$

where $\hat{\hat{Y}}_{ij}$ is defined as in Lemma 6, Chapter 3.

Unbiasedness of ${}_1\bar{Y}_2$: From Lemma 6

$$ E\,{}_1\hat{\hat{Y}}_2 = \bar{\bar{Y}}_2 \quad . $$

Variance of ${}_1\bar{Y}_2$:

$$ V({}_1\hat{\hat{Y}}_2) = a^2\ V(\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) + c^2\ V(\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) + V(\hat{\hat{Y}}_{22}) $$
$$ + 2ac\ \mathrm{cov}\ (\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}, \hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) + 2a\ \mathrm{cov}\ (\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}, \bar{\bar{Y}}_{22}) $$
$$ + 2c\ \mathrm{cov}\ (\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}, \hat{\hat{Y}}_{22}) \quad . $$

For multi-stage sampling, the variances, covariances are as follows:

$$V(\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) = V(\hat{\hat{Y}}_{11}) + V(\hat{\hat{Y}}_{12}) - \text{cov}\,(\hat{\hat{Y}}_{11}\hat{\hat{Y}}_{12})$$

$$= \frac{N}{N-1}\ \frac{\sigma^2_{1,1}}{nPQ} + \frac{\sigma^2_{1,2}}{nPQ}$$

$$\text{cov}\,(\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12},\ \hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) = \text{cov}\,(\hat{\hat{Y}}_{11}\hat{\hat{Y}}_{21}) - \text{cov}\,(\hat{\hat{Y}}_{12}\hat{\hat{Y}}_{21})$$

$$= -\frac{N}{N-1}\ \frac{\sigma_{12,1}}{Pn} - \frac{\sigma_{12,2}}{Pn}$$

$$\text{cov}\,(\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22},\ \hat{\hat{Y}}_{22}) = \text{cov}\,(\hat{\hat{Y}}_{21}\hat{\hat{Y}}_{22}) - V(\hat{\hat{Y}}_{22})$$

$$= -\frac{N}{N-1}\ \frac{\sigma^2_{2,1}}{Qn} - \frac{\sigma^2_{2,2}}{Qn} \quad .$$

Therefore

$$V({}_1\hat{\hat{Y}}_2) = a^2 \left(\frac{N}{N-1}\frac{\sigma^2_{1,1}}{nPQ} + \frac{\sigma^2_{1,2}}{nPQ}\right) + c^2 \left(\frac{N}{N-1}\ \frac{\sigma^2_{2,1}}{nPQ} + \frac{\sigma^2_{2,2}}{nPQ}\right)$$

$$- 2ac \left(\frac{N}{N-1}\frac{\sigma_{12,1}}{Pn} + \frac{\sigma_{12,2}}{Pn}\right) - 2c \left(\frac{N}{N-1}\frac{\sigma^2_{2,1}}{Qn} + \frac{\sigma^2_{2,2}}{Qn}\right)$$

$$+ \frac{\sigma^2_{2,1}}{Qn}\left(\frac{N-Qn}{N-1}\right) + \frac{\sigma^2_{2,2}}{Qn} \qquad (4.19)$$

<u>Optimum a, c:</u> To determine the optimum weights, we proceed as follows:

$$\frac{\partial}{\partial a}\ V({}_1\hat{\hat{Y}}_2) = a\left(\frac{N}{N-1}\frac{\sigma^2_{1,1}}{nPQ} + \frac{\sigma^2_{1,2}}{nPQ}\right) - c\left(\frac{N}{N-1}\ \frac{\sigma_{12,1}}{Pn} + \frac{\sigma_{12,2}}{Pn}\right) = 0$$

$$\frac{\partial}{\partial c}\ V({}_1\hat{\hat{Y}}_2) = -a\left(\frac{N}{N-1}\ \frac{\sigma_{12,1}}{Pn} + \frac{\sigma_{12,2}}{Pn}\right) + c\left(\frac{N}{N-1}\ \frac{\sigma^2_{2,1}}{nPQ}\right) = \frac{N}{N-1}\frac{\sigma^2_{2,1}}{Qn} + \frac{\sigma^2_{2,2}}{Qn} \quad .$$

Let us denote $\bar{\sigma}_1^2 = \sigma^2_{1,1} + \frac{N-1}{N}\,\sigma^2_{1,2}$

$$\bar{\sigma}_{12} = \sigma_{12,1} + \frac{N-1}{N}\,\sigma_{12,2}$$

$$\bar{\sigma}_2^2 = \sigma^2_{2,1} + \frac{N-1}{N}\,\sigma^2_{2,2} \quad .$$

Then, the optimum equations are:

$$a \frac{\bar{\sigma}_1^2}{nPQ} - c \frac{\bar{\sigma}_{12}}{Pn} = 0$$

$$- a \frac{\bar{\sigma}_{12}}{Pn} + c \frac{\bar{\sigma}_2^2}{nPQ} = \frac{\bar{\sigma}_2^2}{Qn}$$

which yields a_w, c_w, the optimum values of a, c

$$a_w = \frac{\bar{\sigma}_{12}\bar{\sigma}_2^2 \; PQ}{\bar{\sigma}_1^2\bar{\sigma}_2^2 - Q^2\bar{\sigma}_{12}^2}$$

$$c_w = \frac{\bar{\sigma}_1^2\bar{\sigma}_2^2 P}{\bar{\sigma}_1^2\bar{\sigma}_2^2 - Q^2\bar{\sigma}_{12}^2} \qquad (4.20)$$

In order to estimate a_w and c_w, let us define the following

$$s_{1.1}^2 = E(\hat{Y}_{1i} - \bar{\hat{Y}}_1)^2 \text{ , where } \hat{Y}_{1i} = \frac{N_i}{n_i} \sum_j Y_{1ij}, \; \bar{\hat{Y}}_1 = \frac{1}{n} \sum_j \frac{N_i}{n_i} \sum_j Y_{1ij}$$

thus $s_1^2 = \frac{N-1}{N} \; s_{1.1}^2$ is an unbiased estimator of $\bar{\sigma}_1^2$, because

$$\frac{N-1}{N}\left[s_{1.1}^2 - \frac{1}{n} \Sigma \, N_i^2 \; \frac{s_{1.2i}^2}{n_i} \left(\frac{N_i - n_i}{N_i}\right)\right]$$

is an unbiased estimator of $\sigma_{1.1}^2$, where

$$s_{1.2i}^2 = \frac{\Sigma(Y_{1ij} - \bar{\hat{Y}}_{ij})^2}{n_i - 1} \quad , \quad \bar{\hat{Y}}_{1i} = \frac{\Sigma Y_{1ij}}{n_i}$$

and

$$s_{1.2}^2 = \frac{1}{n} \Sigma N_i^2 \; \frac{S_{1.2i}^2}{N_i} \left(\frac{N_i - n_i}{N_i}\right) \text{ is an unbiased estimator of } \sigma_{1.2}^2 \; .$$

Similarly,

$$\bar{s}_2^2 = \frac{N-1}{n} \; s_{2.1}^2, \text{ which is an unbiased estimator of } \bar{\sigma}_2^2$$

$$\bar{s}_{12}^2 = \frac{N-1}{N} \; s_{12.1}^2, \text{ which is an unbiased estimator of } \bar{\sigma}_{12} \; .$$

Then, a_w and c_w can be estimated with bias by $\hat{a}_w$ and $\hat{c}_w$;

$$\hat{a}_w = \frac{\bar{\bar{s}}_{12}\ \bar{\bar{s}}_2^2\ PQ}{\bar{\bar{s}}_1^2\bar{\bar{s}}_2^2 - Q^2\bar{\bar{s}}_{12}^2}$$

$$\hat{c}_w = \frac{\bar{\bar{s}}_1^2\bar{\bar{s}}_2^2\ P}{\bar{\bar{s}}_1^2\bar{\bar{s}}_2^2 - Q^2\bar{\bar{s}}_{12}^2} \quad .$$

Hence, the optimum estimator is

$${}_1\hat{\hat{Y}}_w = a_w\,\hat{\hat{Y}}_{11} + b_w\hat{\hat{Y}}_{12} + c_w\hat{\hat{Y}}_{21} + d_w\,\hat{\hat{Y}}_{22}$$

$$\text{where} \quad a_w + b_w = 0 \quad , \quad c_w + d_w = 1 \quad .$$

The variance of this optimum estimator is from (4.23) and (4.24)

$$V({}_1\hat{\hat{Y}}_w) = \frac{\sigma_{2.1}^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + \frac{\sigma_{2.2}^2}{Qn} - \frac{N}{N-1}\,\frac{\bar{\bar{\sigma}}_2^2}{nPQ}\,\frac{\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2P^2}{(\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q^2\bar{\bar{\sigma}}_{12}^2)}$$

$$= \frac{N}{N-1}\,\frac{\bar{\bar{\sigma}}_2^2}{Qn} - \frac{N}{N-1}\,\frac{\bar{\bar{\sigma}}_2^2}{nPQ}\,\frac{\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2P^2}{(\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q^2\bar{\bar{\sigma}}_{12}^2)} - \frac{\sigma_{2.1}^2}{N-1}$$

$$= \frac{N}{N-1}\,\frac{\bar{\bar{\sigma}}_2^2}{n}\,\frac{(\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q\bar{\bar{\sigma}}_{12}^2)}{\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q^2\bar{\bar{\sigma}}_{12}^2)} - \frac{\sigma_{2.1}^2}{N-1} \quad .$$

In order to compare $V({}_1\hat{\hat{Y}}_w)$ with the variance of a sample mean of size n on the second occasion only,

$$V(\hat{\hat{Y}}_2) = \frac{\sigma_{2.1}^2}{N-1}\left(\frac{N-n}{N-1}\right) + \frac{\sigma_{2.2}^2}{n} = \frac{N}{N-1}\,\frac{\bar{\bar{\sigma}}_2^2}{n} - \frac{\sigma_{2.1}^2}{N-1} \quad .$$

Since

$$\frac{\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q\bar{\bar{\sigma}}_{12}^2}{\bar{\bar{\sigma}}_1^2\bar{\bar{\sigma}}_2^2 - Q^2\bar{\bar{\sigma}}_{12}^2} < 1 \quad ,$$

$$V({}_1\hat{\bar{Y}}_w) \text{ is smaller than } V(\hat{\bar{Y}}_2) \quad .$$

The optimum value of Q for given ρ is obtained as a function of $\bar{\sigma}_1$, $\bar{\sigma}_2$, $\bar{\sigma}_{12}$, i.e.,

$$Q_{opt} = \frac{\bar{\sigma}_1^2\bar{\sigma}_2^2 - \bar{\sigma}_1\bar{\sigma}_2\sqrt{\bar{\sigma}_1^2\bar{\sigma}_2^2 - \bar{\sigma}_{12}^2}}{\bar{\sigma}_{12}^2} \quad .$$

4.5. A Discussion of a Symmetric Linear Estimator

So far, we have considered the four classes of estimators of $\bar{\bar{Y}}_2$ for samples on two successive occasions. Those estimators have been studied by others based on sampling with replacement at each draw. In this section, a so called "symmetric linear estimator" will be introduced and discussed. This new estimator will be connected with the theory of estimation for samples on more than two successive occasions.

4.5.1. A General Discussion

The four classes of estimators discussed in the previous sections are linear combinations of $\bar{Y}_{11}$, $\bar{Y}_{12}$, $\bar{Y}_{21}$ and $\bar{Y}_{22}$, and each estimator is defined according to a different set of coefficients. However, if we observe carefully, they have a point in common. Namely, the units in the sample of size n + Qn are partitioned into three sets, and each sample mean is based on one of the three sets. One is the set s_1 which consists of those units belonging to the first occasion only. Second, is the set s_2 of matched units belonging to both occasions. Third, is that s_3 of those units belonging to the second occasion only. Then $\bar{Y}_{11}$ is based on s_1, $\bar{Y}_{12}$ and $\bar{Y}_{21}$ are based on s_2, and $\bar{Y}_{22}$ is based on s_3, as mentioned in

4.2. Each estimator is distinguished only by different coefficients of $\bar{Y}_{(2-i)j}$, as mentioned in 4.3.

It is not necessary, however, to confine the partitioning to three sets. In fact, we can form a large number of linear combination of $\bar{Y}_{(2-i)j}$ by partitioning the first sample arbitrarily into two sets (or more than two) and by partitioning the second sample arbitrarily into two or more sets. The following two estimators will be discussed, among the possible many estimators.

Let us partition the first sample of size n into two sets, where the first Qn units are denoted by the set s_1 and the remaining $n - Qn = Pn$ units are denoted by the set s_2. Also, let us partition the second sample of size n in the same manner as the first sample, denoting the first Qn units in this sample by s_3 and the remaining Pn units by s_4. We note that if all the $n + Qn$ different units in the two samples are numbered serially from 1 to $n + Qn$, then s_2 consists of those numbered $Qn + 1$ to n, s_3 consists of those numbered $Qn + 1$ to $2Qn$ and s_4 consists of those numbered $2Qn + 1$ to $n + Qn$. Figure 4.1 may make this pattern clear. Based on this symmetric partition of a sample, each sample mean $\bar{Y}_{(2-i)j}$ can be computed and a linear combination of these four $\bar{Y}_{(2-i)j}$s is also an estimator of $\bar{\bar{Y}}_2$. Let us call this estimator $_6\bar{\bar{Y}}_2$.

In contrast with $_5\bar{\bar{Y}}_2$, we can construct another linear combination based on the pattern in which Qn, Pn are interchanged as in Figure 4.2. Let us call this $\bar{\bar{Y}}_2$. Since both $_5\bar{\bar{Y}}_2$ and $_6\bar{\bar{Y}}_2$ are based on a symmetric pattern of partitioning of a sample, we call them symmetric estimators.

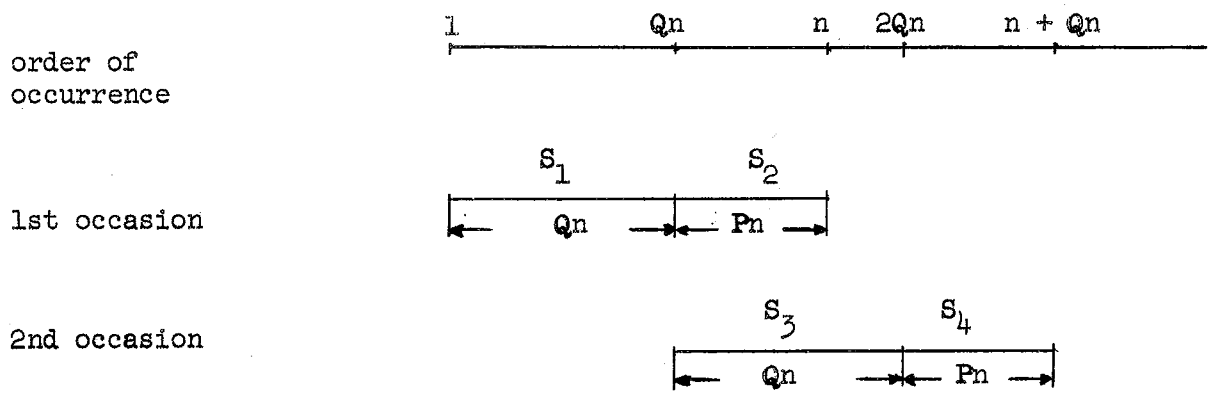


Figure 4.1. A symmetric pattern for ${}_5\bar{Y}_2$

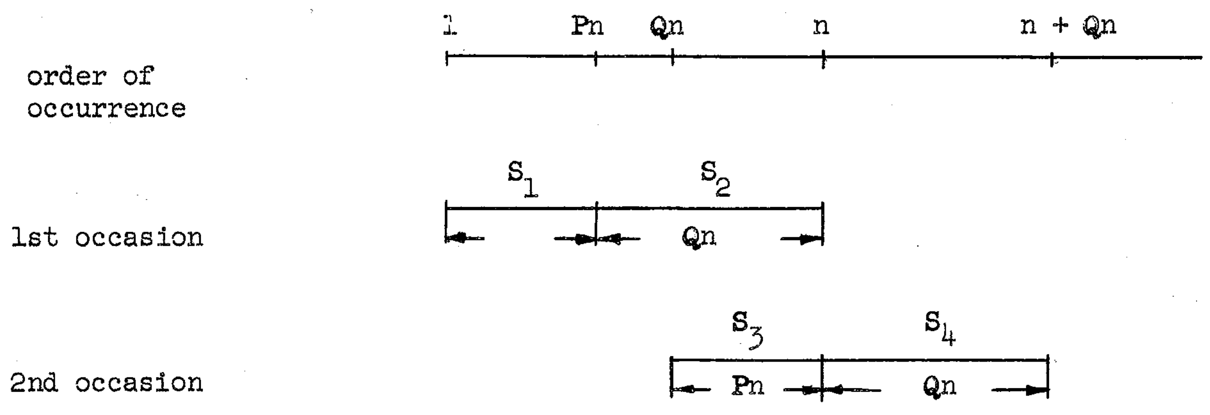


Figure 4.2. A symmetric pattern for ${}_6\bar{Y}_2$

We can see the symmetric patterns for ${}_5\bar{Y}_2$, ${}_6\bar{Y}_2$ are different from that for ${}_1\bar{Y}_2$ (or the other three). The estimates ${}_5\bar{Y}_2$ and ${}_6\bar{Y}_2$ are the same as the general linear estimator ${}_1\bar{Y}_2$ only when P and Q are equal. Since we are interested in a symmetric estimator in developing the theory of estimation for more than two occasions, the properties of ${}_5\bar{Y}_2$ and ${}_6\bar{Y}_2$ will be investigated in the next step.

Since the optimum value of Q is always larger than .5, the following analysis will be confined to values of Q larger than .5.

4.5.2. Symmetric Linear Estimators and Their Efficiency

1. A symmetric linear estimator ${}_5\bar{Y}_2$

As mentioned previously, a symmetric linear estimator ${}_5\bar{Y}_2$ is a linear combination of sample means $\bar{Y}_{(2-i)j}$ which are based on the partitioning pattern described in 4.5.1.

$$ {}_5\bar{Y}_2 = a\,\bar{Y}_{11} + b\bar{Y}_{12} + c\,\bar{Y}_{21} + d\,\bar{Y}_{22} $$

$$ \text{where} \quad a + b = 0 \qquad c + d = 1 $$

and ${}_5\bar{Y}_2$ is an unbiased estimator of $\bar{\bar{Y}}_2$.

The variance of ${}_5\bar{Y}_2$: ${}_5\bar{Y}_2$ can be rewritten as

$$ {}_5\bar{Y}_2 = a\bar{Y}_{11} + b\bar{Y}_{12} + c\bar{Y}_{21} + d\bar{Y}_{22} = a(\bar{Y}_{11} - \bar{Y}_{12}) + c(\bar{Y}_{21} - \bar{Y}_{22}) + \bar{Y}_{22} $$

$$ V({}_5\bar{Y}_2) = a^2 V(\bar{Y}_{11} - \bar{Y}_{12}) + c^2 V(\bar{Y}_{21} - \bar{Y}_{22}) + V(\bar{Y}_{22}) $$

$$ + 2ca \operatorname{cov}(\bar{Y}_{11} - \bar{Y}_{12}, \bar{Y}_{21} - \bar{Y}_{22}) + 2c\,(\bar{Y}_{21} - \bar{Y}_{22}, \bar{Y}_{22}) \; , $$

where $$V(\bar{Y}_{11} - \bar{Y}_{12}) = \frac{N}{N-1} \frac{\sigma_1^2}{nPQ}$$

$$\text{cov}\,(\bar{Y}_{11} - \bar{Y}_{12}, \bar{Y}_{21} - \bar{Y}_{22}) = -\frac{N}{N-1} \frac{\sigma_{12}}{Q_n}$$

$$\text{cov}\,(\bar{Y}_{21} - \bar{Y}_{22}, \bar{Y}_{22}) = -\frac{N}{N-1} \frac{\sigma_2^2}{P_n}$$

<u>Optimum estimator and its variance for given Q and ρ</u>

$$\frac{(N-1)}{Z} \frac{nPQ}{N} \frac{\partial}{\partial a} V({}_5\bar{Y}_2) = a\sigma_1^2 - cP\sigma_{12} = 0$$

$$\frac{(N-1)}{2N} \frac{nPQ}{N} \frac{\partial}{\partial c} V({}_5\bar{Y}_2) = -aP\sigma_{12} + c\sigma_2^2 - Q\sigma_2^2 = 0 \quad .$$

In matrix form, $A_2\underline{P} = \underline{q}$

$$A_2 = \begin{pmatrix} \sigma_1^2 & -P\sigma_{12} \\ -P\sigma_{12} & \sigma_2^2 \end{pmatrix} \qquad \underline{P} = \begin{pmatrix} a \\ c \end{pmatrix} \qquad \underline{q} = \begin{pmatrix} 0 \\ Q\sigma_2^2 \end{pmatrix}$$

$$|A_2| = \sigma_1^2\sigma_2^2\,(1 - P^2\rho^2) \quad .$$

Solving for a and c, we have

$$a_w = \frac{\rho PQ}{1-P^2\rho_{12}^2} \frac{\sigma_2}{\sigma_1}$$

$$c_w = \frac{Q}{1-P^2\rho_{12}^2} \quad .$$

Hence, the optimum estimator of $\bar{\bar{Y}}_2$ in this class is

$${}_5\bar{Y}_w = a_w\,(\bar{Y}_{11} - \bar{Y}_{12}) + c_w(\bar{Y}_{21} - \bar{Y}_{22}) + \bar{Y}_{22}$$

and its variance is

$$V({}_5\bar{Y}_w) = \frac{\rho^2 P^2 Q^2}{(1-P^2\rho_{12}^2)} \frac{\sigma_2^2}{\sigma_1^2} \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} + \frac{Q^2}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} + \frac{\sigma_2^2}{pn} \left(\frac{N-Pn}{N-1}\right)$$

$$- 2 \frac{\rho^2 P^2 Q^2}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} - 2 \frac{Q^2}{1-P^2\rho^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ}$$

$$= \frac{\sigma_2^2}{Pn} \left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} \frac{Q^2}{(1-P^2\rho^2)} . \qquad (4.21)$$

This is the variance of the optimum estimator ${}_5\bar{Y}_w$ for a given value of ρ when the sampling is carried out according to the proposed scheme described in 3.1.

2. A symmetric linear estimator ${}_6\bar{Y}_2$

This estimator is also formed according to the partitioning pattern described in 4.5.1

$${}_6\bar{Y}_2 = a\bar{Y}_{11} + b\bar{Y}_{12} + c\bar{Y}_{21} + d\bar{Y}_{22}$$

$$\text{where} \quad a + b = 0 \qquad c + d = 1 .$$

This is also an unbiased estimator of $\bar{\bar{Y}}_2$.

<u>The variance of ${}_6\bar{Y}_2$:</u>

$$V({}_6\bar{Y}_2) = a^2 V(\bar{Y}_{11} - \bar{Y}_{12}) + c^2 V(\bar{Y}_{21} - \bar{Y}_{22}) + V(\bar{Y}_{22})$$

$$+ 2ac \operatorname{cov}(\bar{Y}_{11} - \bar{Y}_{12}, \bar{Y}_{21} - \bar{Y}_{22}) + 2c \operatorname{cov}(\bar{Y}_{21} - \bar{Y}_{22}, \bar{Y}_{22})$$

where $\operatorname{cov}(\bar{Y}_{11} - \bar{Y}_{12}, \bar{Y}_{21} - \bar{Y}_{22}) = \operatorname{cov}(\bar{Y}_{11}\bar{Y}_{21}) - \operatorname{cov}(\bar{Y}_{12}\bar{Y}_{21})$

$$= -\frac{\sigma_{12}}{N-1} - \frac{\sigma_{12}}{Qn}\left(\frac{N-Qn}{N-1}\right) = -\frac{N}{N-1} \frac{\sigma_{12}}{Qn}$$

$$\operatorname{cov}(\bar{Y}_{21} - \bar{Y}_{22}, \bar{Y}_{22}) = -\frac{N}{N-1} \frac{\sigma_2^2}{Qn} .$$

<u>Optimum estimator and its variance for given Q and ρ:</u>

$$\frac{(N-1)nPQ}{2N} \frac{\partial}{\partial a} V({}_6\bar{\bar{Y}}_2) = a\sigma_1^2 - cP\sigma_{12} = 0$$

$$\frac{(N-1)nPQ}{2N} \frac{\partial}{\partial c} V({}_6\bar{\bar{Y}}_2) = -aP\sigma_{12} + c\sigma_2^2 = P\sigma_2^2 \quad .$$

In matrix form $A_2\underline{P} = \underline{q}$

$$A_2 = \begin{bmatrix} \sigma_1^2 & -P\sigma_{12} \\ -P\sigma_{12} & \sigma_2^2 \end{bmatrix} \qquad \underline{P} = \begin{bmatrix} a \\ c \end{bmatrix} \qquad \underline{q} = \begin{bmatrix} 0 \\ PQ_2^2 \end{bmatrix} \quad .$$

Solving for a and c, we have

$$a_w = \frac{\rho P^2}{1-P^2\rho^2} \frac{\sigma_2}{\sigma_1}$$

$$c_w = \frac{P}{1-P^2\rho^2} \quad .$$

The optimum estimator of $\bar{\bar{Y}}_2$ in this class is

$${}_6\bar{Y}_w = a_w(\bar{Y}_{11} - \bar{Y}_{12}) + c_w(\bar{Y}_{21} - \bar{Y}_{22}) + \bar{Y}_{22}$$

and its variance is

$$V({}_6\bar{Y}_w) = \frac{\rho^2P^4}{(1-P^2\rho^2)^2} \frac{\sigma_2^2}{\sigma_1^2} \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} + \frac{P^2}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} + \frac{\sigma_2^2}{Qn} \left(\frac{N-Qn}{N-1}\right)$$

$$- 2\frac{\rho^2P^4}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} - 2\frac{P^2}{(1-P^2\rho^2)} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ}$$

$$= \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1} \frac{P^2}{(1-P^2\rho^2)} \frac{\sigma_2^2}{nPQ} \qquad (4.22)$$

$$= \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} \left(P - \frac{P^2}{1-P^2\rho^2}\right) \quad . \qquad (4.23)$$

3. Efficiency of the symmetric estimators.

Let us consider efficiency of the estimators ${}_5\overline{Y}_2$ and ${}_6\overline{Y}_2$.

${}_1\overline{Y}_2$ versus ${}_5\overline{Y}_2$: From (4.2) and (4.21), we have

$$V({}_1\overline{Y}_w) = \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\frac{P^2}{1-P^2\rho^2} = \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\left(P - \frac{P^2}{1-Q^2\rho^2}\right) - \frac{\sigma_2^2}{N-1} \qquad (4.24)$$

$$V({}_5\overline{Y}_w) = \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\frac{Q^2}{1-P^2\rho^2} = \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\left(Q - \frac{Q^2}{1-P^2\rho^2}\right) - \frac{\sigma_2^2}{N-1}. \qquad (4.25)$$

Therefore, ${}_1\overline{Y}_2$ is better than ${}_5\overline{Y}_2$ if

$$P - \frac{P^2}{1-Q^2\rho^2} > Q - \frac{Q^2}{1-P^2\rho^2}$$

which implies

$$Q(1-P^2\rho^2-Q)(1-Q^2\rho^2) > P(1-Q^2\rho^2-P)(1-P^2\rho^2)$$

$$0 > PQ\rho^4(P-Q) \quad .$$

This is always true when $Q > P$. Therefore, ${}_1\overline{Y}_2$ has smaller variance than ${}_5\overline{Y}_2$ when $Q > P$.

${}_1\overline{Y}_2$ versus ${}_6\overline{Y}_2$: From (4.2), (4.22), we have

$$V({}_6\overline{Y}_w) = \frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\frac{P^2}{(1-P^2\rho^2)} \quad .$$

Therefore, ${}_1\overline{Y}_2$ has smaller variance than ${}_6\overline{Y}_2$ if

$$\frac{P^2}{1-Q^2\rho^2} > \frac{P^2}{1-P^2\rho^2} \quad ,$$

which implies $1 - Q^2\rho^2 < 1 - P^2\rho^2$, $\rho^2(P-Q) < 0$.

This is true when $Q > P$.

$_5\bar{\bar{Y}}_2$ versus $_6\bar{\bar{Y}}_2$: Now we consider the efficiency between the symmetric estimators. From (4.23) and (4.25), $_5\bar{\bar{Y}}_2$ is better than $_6\bar{\bar{Y}}_2$ if

$$P - \frac{P^2}{1-P^2\rho^2} > Q - \frac{Q^2}{1-P^2\rho^2} \quad .$$

This yields $0 > P\rho^2(P-Q)$ which is true if $Q > P$. Therefore, $_5\bar{\bar{Y}}_2$ has smaller variance $_6\bar{\bar{Y}}_2$ when $Q > P$. In conclusion, $_1\bar{\bar{Y}}_2$ is uniformly better in the least variance sense than $_5\bar{\bar{Y}}_2$, $_6\bar{\bar{Y}}_2$ for all values of ρ and $Q > P$. The important fact is that the general linear estimator based on a complete matching pattern has smaller variance than a general linear estimator based on an incomplete matching pattern for estimating $\bar{\bar{\bar{Y}}}_2$ on successive occasions. Also $_5\bar{\bar{Y}}_2$ is better than $_6\bar{\bar{Y}}_2$ when $Q > P$. Because of the symmetric formation, it seems that both estimators are equivalent. However, we observe that the efficiency depends on the Q value.

Table 4.11 gives a summary of our analysis. Relative precision is defined as

$$R \cdot P(\%) = \frac{V(_1\bar{\bar{Y}}_2)}{V(_1\bar{\bar{Y}}_2)} \cdot 100$$

To avoid complexity, we ignore the $\frac{\sigma_2^2}{N-1}$ term in each variance.

Table 4.11. Precision of the estimators $_2\bar{\bar{Y}}_2$, $_3\bar{\bar{Y}}_2$, $_5\bar{\bar{Y}}_2$ and $_6\bar{\bar{Y}}_2$ relative to $_1\bar{\bar{Y}}_2$

Q	.6				.7				.8				.9			
ρ \ $_i\bar{\bar{Y}}_2$	$_2\bar{\bar{Y}}_2$	$_3\bar{\bar{Y}}_3$	$_5\bar{\bar{Y}}_2$	$_6\bar{\bar{Y}}_2$	$_2\bar{\bar{Y}}_2$	$_3\bar{\bar{Y}}_2$	$_5\bar{\bar{Y}}_2$	$_6\bar{\bar{Y}}_2$	$_2\bar{\bar{Y}}_2$	$_3\bar{\bar{Y}}_2$	$_5\bar{\bar{Y}}_2$	$_6\bar{\bar{Y}}_2$	$_2\bar{\bar{Y}}_2$	$_3\bar{\bar{Y}}_2$	$_5\bar{\bar{Y}}_2$	$_6\bar{\bar{Y}}_2$
.5	.94	.98	1.00	.96	.94	.98	1.00	.95	.95	.99	.99	.96	.97	.99	.99	.97
.6	.95	.97	.99	.94	.95	.97	.99	.92	.96	.98	.98	.93	.98	.987	.99	.96
.7	.97	.95	.98	.91	.97	.96	.97	.88	.97	.97	.96	.90	.98	.985	.97	.93
.8	.98	.94	.97	.87	.98	.95	.94	.83	.98	.96	.93	.87	.98	.98	.94	.88
.9	.99	.92	.93	.81	.99	.93	.88	.74	.99	.95	.85	.84	.99	.98	.85	.79

Figure 4.3 shows the graph of Table 4.11. From Figure 4.3, we may point out that when $Q > P$, ${}_2\bar{Y}_2$ is superior for high values of correlation (e.g., $\rho > .75$), ${}_5\bar{Y}_2$ is superior for low values of correlation ($\rho < .75$) and ${}_3\bar{Y}_2$ is dominated by either ${}_2\bar{Y}_2$ or ${}_5\bar{Y}_2$. The worst in the domain of $Q > .5$ is ${}_6\bar{Y}_2$.

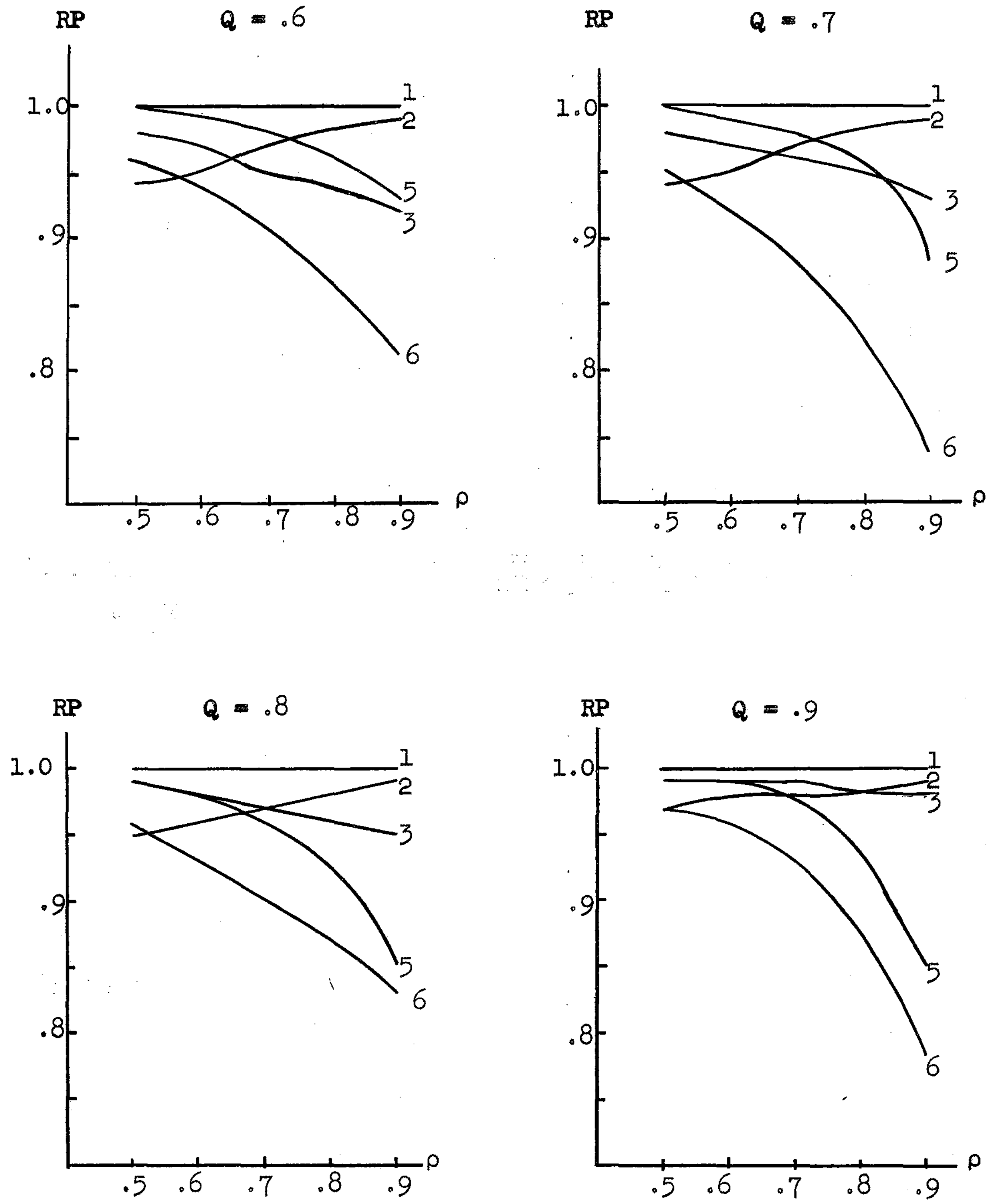


Figure 4.3. Precision of the estimators $_2\bar{Y}_2$, $_3\bar{Y}_2$, $_5\bar{Y}_2$ and $_6\bar{Y}_2$ relative to $_1\bar{Y}_2$

5. ESTIMATION THEORY FOR THE SAMPLING ON MORE THAN TWO SUCCESSIVE OCCASIONS

5.1. Introductory Remarks

We have seen, in Chapter 4, the estimation theory for two occasions based on the proposed sampling scheme in 3.1. The conclusion obtained there is that the general linear estimator is the best among all estimators under consideration in this thesis. Accordingly, it is natural to develop the general linear estimator for sampling on more than two occasions based on the proposed sampling scheme.

A general linear estimator for more than two successive occasions will be defined as follows: Suppose the total of $n + (\alpha-1)Qn$ units in the sample for α occasions is partitioned into disjoint sets such that each unit in a set appears in the sample together with the other units in the same set. Applying this partitioning method to the sample under the proposed sampling scheme, there are $\alpha+(Q-1)$ disjoint sets if $\frac{1}{Q}$ is an integer or $[\frac{1}{Q}] + 1 + 2(\alpha-1)$ disjoint sets if $\frac{1}{Q}$ is a fractional number, where $[\frac{1}{Q}]$ is the largest integer less than $\frac{1}{Q}$. Thus the general linear estimator as defined here is a linear combination of $\bar{Y}_{(\alpha-i)j}$ $(i = 0,\ldots, \alpha-1 \quad , \quad 1 \leq j \leq \frac{1}{Q})$ which is the sample mean of the j^{th} set on occasion α-i.

$$_1\bar{Y}_\alpha = \sum_i \sum_j a_{\alpha-i}\bar{Y}_{(\alpha-i)j}$$

$$\text{where} \quad \sum_j a_{(\alpha-i)j} = 0 \quad \text{when} \quad i \neq 0$$
$$= 1 \quad \text{when} \quad i = 0$$

It is suspected that, with the proposed sampling scheme, this estimator has smaller variance than any other linear unbiased estimator. This contention will be demonstrated by an example later.

However, the process of finding the minimum variance unbiased estimator which belongs to this class is lengthy and complicated especially when **Q** is small. The estimation process becomes simpler when the sample on each occasion is partitioned into just two sets, thus involving only a small number of coefficients. In fact, the symmetric estimator in 4.5 is this type of estimator.

Therefore, as in the discussion in 4.5, we will study the symmetric estimator and the modified composite estimator for more than two occasions. Actually, the symmetric estimator is also a kind of a general linear estimator, based on a symmetric pattern of partitioning a sample, and its efficiency was fairly good for $.5 < \rho < .75$ in the two occasions case. The efficiency of the modified composite estimator was also good for the values of ρ larger than .75. In 5.2 and 5.3, some properties of the symmetric estimator and the modified composite estimator will be investigated, and then comparison between these estimators will be carried out in 5.4. Considering that the optimum values of **Q** on two successive occasions in 4.2 are always larger than .5, the estimation theory for both estimators treated in this chapter is confined to the region of **Q** larger than .5. The numerical analysis in the following chapters was performed by an IBM 1620 computer.

5.2. A Symmetric Estimator and Its Properties

A symmetric estimator on the α^{th} occasion ${}_5\bar{Y}_\alpha$, of $\bar{\bar{Y}}_\alpha$, as an extension of ${}_5\bar{Y}_2$ in 4.5, is defined as a linear combination of the sample means $\bar{Y}_{(\alpha-i)j}$ $(i = 0, \ldots, \alpha-1, j=1,2)$, where $\bar{Y}_{(\alpha-i)1}$ is the sample mean of

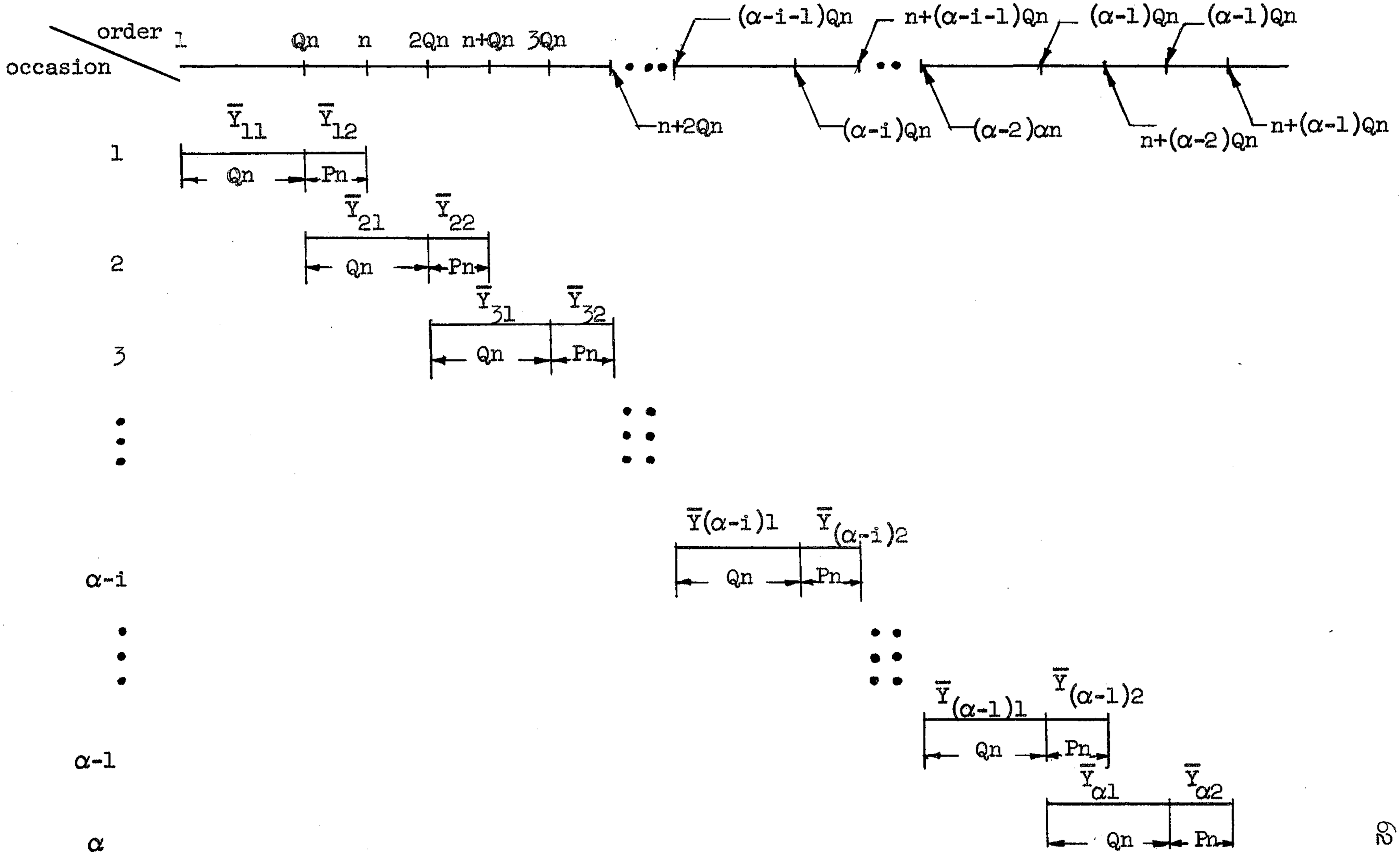


Figure 5.1. Formation of $_5\bar{Y}_\alpha$.

Qn units appearing from order $(\alpha-i-1)Qn + 1$ to $(\alpha-i)Qn$ on the $(\alpha-i)$th occasion and $\bar{Y}_{(\alpha-i)2}$ is the same mean of Pn units appearing from order $(\alpha-i)Qn+1$ to $n+(\alpha-i-1)Qn$ on the $(\alpha-i)$th occasion. Figure 5.1 shows the formation of ${}_5\bar{Y}_{\alpha.}$.

Let i denote the number of previous occasions used in estimating the population mean $\bar{\bar{Y}}_\alpha$ on the αth occasion. For each i there is an estimator of $\bar{\bar{Y}}_\alpha$. Let us denote this sequence of estimators by $\{{}_5\bar{Y}_{\alpha_i}\}$ and ${}_5\bar{Y}_{\alpha_{\alpha-i}} = {}_5\bar{Y}_{\alpha.}$. Then, an estimator of $\bar{\bar{Y}}_\alpha$, using all $\alpha-1$ previous occasions, can be written

$$ {}_5\bar{Y}_{\alpha.} = \sum_{i=0}^{\alpha-1} a_i \bar{Y}_{(\alpha-i)1} + \sum_{i=0}^{\alpha-1} b_i \bar{Y}_{(\alpha-i)2} = \bar{Y}_{\alpha 2} + \sum_{i=0}^{\alpha-1} a_i (\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}) , $$

where $a_o + b_o = 1$, $a_i + b_o = 0$, $i \neq 0$.

<u>Unbiasedness of ${}_5\bar{Y}_\alpha$:</u> From Lemma 1, ${}_5\bar{Y}_{\alpha.}$ is an unbiased estimator of the population mean on the αth occasion, i.e.,

$$ E\ {}_5\bar{Y}_{\alpha.} = \bar{\bar{Y}}_\alpha \quad . $$

<u>Variance of ${}_5\bar{Y}_\alpha$:</u> The variance of ${}_5\bar{Y}_{\alpha.}$ is

$$ V({}_5\bar{Y}_{\alpha.}) = V(\bar{Y}_{\alpha 2}) + \sum_{i=0}^{\alpha-1} a_i^2 V(\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2} $$

$$ + 2 \sum_{i=0}^{\alpha-1} a_i \text{ cov } (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}) + 2 \sum_{i<j} a_i a_j $$

$$ \text{cov}(\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}, \bar{Y}_{(\alpha-j)1} - \bar{Y}_{(\alpha-j)2}) \ . $$

From the lemmas in 3.2, each variance and covariance in $V({}_5\bar{Y}_\alpha)$ is as follows:

$$ V(\bar{Y}_{\alpha 2}) = \frac{\sigma_\alpha^2}{Pn} = (\frac{N-Pn}{N-1}) . \quad V(\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}) = V(\bar{Y}_{(\alpha-i)1}) + V(\bar{Y}_{(\alpha-i)2}) $$

$$ - 2\text{cov } (\bar{Y}_{(\alpha-i)1}, \bar{Y}_{(\alpha-i)2}) = \frac{\sigma_{\alpha-i}^2}{Qn} (\frac{N-Qn}{N-1}) + \frac{\sigma_{\alpha-i}^2}{Pn} (\frac{N-Pn}{N-1}) + 2 \frac{\sigma_{\alpha-i}^2}{N-1} = \frac{N}{N-1} \frac{\sigma_{\alpha-i}^2}{nPQ} $$

$$\text{Cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{\alpha 1} - \bar{Y}_{\alpha 2}) = \text{cov}\ (\bar{Y}_{\alpha 2}\bar{Y}_{\alpha 1}) - V(\bar{Y}_{\alpha 2}) = -\frac{\sigma_\alpha^2}{N-1} - \frac{\sigma_\alpha^2}{Pn}\left(\frac{N-Pn}{N-1}\right)$$

$$= -\frac{N}{N-1}\ \frac{\sigma_\alpha^2}{Pn}$$

$$\text{Cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}) = \text{cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-i)1}) - \text{cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-i)2}) = 0$$

for all $i = 1, 2, \ldots, \alpha-1$

$$\text{Cov}\ (\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2})(\bar{Y}_{(\alpha-j)1} - \bar{Y}_{(\alpha-j)2}) = \text{cov}\ (\bar{Y}_{(\alpha-i)1}, \bar{Y}_{(\alpha-j)1})$$

$$- \text{cov}\ (\bar{Y}_{(\alpha-i)2}, \bar{Y}_{(\alpha-j)1}) - \text{cov}\ (\bar{Y}_{(\alpha-j)1}, \bar{Y}_{(\alpha-j)2})$$

$$+ \text{cov}\ (\bar{Y}_{(\alpha-i)2}, \bar{Y}_{(\alpha-j)2})$$

$$= -\frac{N}{N-1}\ \frac{\sigma_{(\alpha-i)(\alpha-j)}}{Qn} \qquad \text{if} \quad j = i+1$$

$$= 0 \qquad \text{otherwise} \quad .$$

Since, $\text{cov}\ (\bar{Y}_{\alpha 1} - \bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-1)1} - \bar{Y}_{(\alpha-1)2}) = \text{cov}\ (\bar{Y}_{\alpha 1}, \bar{Y}_{(\alpha-1)1})$

$$- \text{cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-1)1}) - \text{cov}\ (\bar{Y}_{\alpha 1}, \bar{Y}_{(\alpha-1)2})$$

$$+ \text{cov}\ (\bar{Y}_{\alpha 2}, \bar{Y}_{(\alpha-1)2}) = -\frac{\sigma_{\alpha(\alpha-1)}}{N-1} + \frac{\sigma_{\alpha(\alpha-1)}}{N-1}$$

$$- \text{cov}\ (\bar{Y}_{\alpha 1}\bar{Y}_{(\alpha-1)2}) - \frac{\sigma_{\alpha(\alpha-1)}}{N-1}$$

where, $\text{cov}\ (\bar{Y}_{\alpha 1}, \bar{Y}_{(\alpha-1)2}) = \text{cov}\ \left(\frac{\Sigma Y_\alpha}{Qn}, \frac{\Sigma Y_{\alpha-1}}{Pn}\right)$

$$= \text{cov}\ \left(\frac{P}{Q}\bar{Y}_{\alpha 11} + \frac{(Q-P)}{Q}\bar{Y}_{\alpha 12}, \bar{Y}_{(\alpha-1)2}\right)$$

$$= \frac{P}{Q}\ \frac{\sigma_{\alpha(\alpha-1)}}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{Q-P}{Q}\ \frac{\sigma_{\alpha(\alpha-1)}}{N-1} = \frac{N-Qn}{N-1}\ \frac{\sigma_{\alpha(\alpha-1)}}{Qn} \quad .$$

Where, $\overline{Y}_{\alpha 11}$ is the mean of those units appearing from $(\alpha-1)Qn+1$ to $n(\alpha-2)Qn$ on the α^{th} occasion, which is the matched portion with the $\alpha-1$ occasion. Also, $\overline{Y}_{\alpha 12}$ is the mean of those units appearing from $n+(\alpha-2)Qn+1$ to $(\alpha-1)Qn$ on the αth occasion, which is the unmatched portion. Therefore, the variance of ${}_5\overline{Y}_{\alpha.}$ is

$$V({}_5\overline{Y}_{\alpha.}) = \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) + \sum_{i=o}^{\alpha-1} a_i^2 \frac{N}{N-1}\frac{\sigma^2_{\alpha-i}}{nPQ}$$

$$- 2a_o \frac{N}{N-1}\frac{\sigma^2_\alpha}{Pn} - 2\sum a_i a_{i+1} \frac{N}{N-1}\frac{\sigma_{(\alpha-i)(\alpha-i-1)}}{Qn} \quad .$$

Optimum estimator and its variance: The optimum value of a_i can be obtained as usual

$$\frac{\partial V}{\partial a_o} = 2\, a_o \frac{N}{N-1}\frac{\sigma^2_\alpha}{nPQ} - 2\frac{N}{N-1}\frac{\sigma^2_\alpha}{Pn} - 2\, a_1 \frac{N}{N-1}\frac{\sigma_{\alpha(\alpha-1)}}{Qn} = 0$$

$$\frac{\partial V}{\partial a_i} = 2\, a_i \frac{N}{N-1}\frac{\sigma^2_{\alpha-i}}{nPQ} - 2a_{i-1}\frac{N}{N-1}\frac{\sigma_{(\alpha-i+1)(\alpha-i)}}{Qn}$$

$$- 2\, a_{i+1}\frac{N}{N-1}\frac{\sigma^2_{(\alpha-i)(\alpha-i-1)}}{Qn} = 0 \quad , \quad i = 0, 1, \ldots, \alpha-2$$

$$\frac{\partial V}{\partial a_{\alpha-1}} = 2a_{\alpha-1}\frac{N}{N-1}\frac{\sigma^2_1}{nPQ} - 2\, a_{\alpha-2}\frac{N}{N-1}\frac{\sigma_{12}}{Qn} = 0 \quad ,$$

which can be written in matrix form

$$A\,\underline{a} = \underline{c} \quad ,$$

where

$$\underline{a}' = (a_{\alpha-1}, a_{\alpha-2}, \ldots, a_o)$$

$$\underline{c}' = (0, 0, \ldots, Q\sigma^2_\alpha)$$

$$
A_\alpha = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array}
\begin{array}{cccccccc}
a_{\alpha-1} & a_{\alpha-2} & a_{\alpha-3} & a_{\alpha-4} & \cdots & a_2 & a_1 & a_o \\
\end{array}
$$

$$
A_\alpha = \left(\begin{array}{cccccccc}
\sigma_1^2 & -P\sigma_{12} & 0 & 0 & \cdots & 0 & 0 & 0 \\
-P\sigma_{12} & \sigma_2^2 & -P\sigma_{23} & 0 & \cdots & 0 & 0 & 0 \\
0 & -P\sigma_{23} & p_3^2 & -P\sigma_{23} & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots\ \vdots & 0 & 0 & 0 \\
0 & 0 & \cdot & \cdot & -P\sigma_{(\alpha-2)(\alpha-3)} & \sigma_{\alpha-2}^2 & -P\sigma_{(\alpha-1)(\alpha-2)} & 0 \\
0 & 0 & \cdot & \cdot & 0 & -P\sigma_{(\alpha-1)(\alpha-2)} & \sigma_{\alpha-1}^2 & -P\sigma_{\alpha(\alpha-1)} \\
0 & 0 & \cdot & \cdot & 0 & 0 & -P\sigma_{\alpha(\alpha-1)} & \sigma_\alpha^2
\end{array}\right)
\begin{array}{c} a_{\alpha-1} \\ a_{\alpha-2} \\ a_{\alpha-3} \\ \vdots \\ \vdots \\ a_2 \\ a_1 \\ a_o \end{array}
$$

Since A is nonsingular, A has a unique inverse. The solution vector $\underline{a}$ in $A\underline{a} = \underline{c}$ requires only the last column $\underline{d}_\alpha$ in $A^{-1} = (\underline{d}_1\ \underline{d}_2 \ldots d_\alpha)$. In order to find $\underline{d}_\alpha$ and $\underline{a} = A^{-1}\underline{c}$, let us start from i = 1.

(i=1). When the number of previous occasions is one, the matrix A_2 is

$$
A_2 = \begin{pmatrix} \sigma_{\alpha-1}^2 & -P\sigma_{\alpha(\alpha-1)} \\ -P\sigma_{\alpha(\alpha-1)} & \sigma_\alpha^2 \end{pmatrix} .
$$

By a well known theorem $A^{-1} = \dfrac{\text{adj } A}{|A|}$, and the last column $\underline{d}_2 = \begin{pmatrix} d_{21} \\ d_{22} \end{pmatrix}$ in A_2^{-1} is

$$
d_{21} = \frac{P\sigma_{\alpha(\alpha-1)}}{|A_2|}
$$

$$
d_{22} = \frac{\sigma_{\alpha-1}^2}{|A_2|}
$$

where $|A_2| = \sigma_\alpha^2\sigma_{\alpha-1}^2\,(1 - P^2\rho_{\alpha(\alpha-1)}^2) = \sigma_\alpha^2\sigma_{\alpha-1}^2\Delta_2$,

$$\Delta_2 = \begin{vmatrix} 1 & -P\rho_{(\alpha-1)\alpha} \\ -P\rho_{(\alpha-1)\alpha} & 1 \end{vmatrix} = 1 - P^2\rho^2_{\alpha(\alpha-1)} \quad .$$

Since $\underline{a} = \begin{bmatrix} a_1 \\ a_o \end{bmatrix} = \underline{a}_2 Q\sigma^2_\alpha$, we have

$$a_{1w} = d_{21}Q\sigma^2_\alpha = PQ\sigma_{\alpha(\alpha-1)} \frac{\sigma^2_\alpha}{|A_2|} = \rho_{\alpha(\alpha-1)} \frac{\sigma_\alpha}{\sigma_{\alpha-1}} \frac{PQ}{(1-P^2\rho^2_{\alpha(\alpha-1)})}$$

$$a_{ow} = d_{22}Q\sigma^2_\alpha = Q\ \sigma^2_{\alpha-1} \frac{\sigma^2_\alpha}{|A_2|} = \frac{Q}{(1-P^2\rho^2_{(\alpha-1)\alpha})} \quad . \tag{5.1}$$

Hence, the optimum estimator $_5\bar{\bar{Y}}_{\alpha_1 w}$ and its variance are

$$_5\bar{\bar{Y}}_{\alpha_1 w} = a_{ow}\bar{Y}_{\alpha 1} + b_o w\ \bar{Y}_{\alpha 2} + a_{1w}\bar{Y}_{(\alpha-1)1} + b_{1w}\bar{Y}_{(\alpha-2)2} \tag{5.2}$$

$$V(_5\bar{\bar{Y}}_{\alpha_1 w}) = \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) + \frac{Q^2}{1-P^2\rho^2_{(\alpha-1)\alpha}} \frac{N}{N-1} \frac{\sigma^2_\alpha}{nPQ} + \rho^2_{(\alpha-1)\alpha} \frac{\sigma^2_2}{\sigma^2_{\alpha-1}} \frac{P^2Q^2}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma^2_{\alpha-1}}{nPQ}$$

$$- 2\frac{Q^2}{(1-P^2\rho^2)} \frac{N}{N-1} \frac{\sigma^2_\alpha}{nPQ} - 2\rho^2 \frac{\sigma^2_\alpha}{\sigma_{\alpha-1}} \frac{P^2Q^2}{(1-P^2\rho^2)^2} \frac{N}{N-1} \frac{\sigma_{\alpha(\alpha-1)}}{nPQ}$$

$$= \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1} \frac{\sigma^2_\alpha}{nPQ} \frac{Q^2}{(1-P^2\rho^2_{\alpha(\alpha-1)})} \tag{5.3}$$

Letting $_5\phi_1 = \dfrac{1}{1-P^2\rho^2_{\alpha(\alpha-1)}}$, we have

$$V(_5\bar{\bar{Y}}_{\alpha_1 w}) = \frac{N}{N-1} \frac{\sigma^2_\alpha}{nPQ} (Q - Q^2\ _5\phi_1) - \frac{\sigma^2_\alpha}{N-1} \quad .$$

$(i = 2)$

when $i = 2$, the matrix A_3 is

$$A_3 = \begin{bmatrix} \sigma^2_{\alpha-2} & -P\sigma_{(\alpha-2)(\alpha-1)} & 0 \\ -P\sigma_{(\alpha-2)(\alpha-1)} & & -P\sigma_{(\alpha-1)\alpha} \\ 0 & -P\sigma_{(\alpha-1)\alpha} & \sigma^2_{\alpha} \end{bmatrix} .$$

Also, the elements of $\underline{d}_3$ in A_3^{-1} are

$$d_{13} = \frac{1}{|A_3|} P^2 \sigma_{(\alpha-2)(\alpha-1)} \sigma_{(\alpha-1)\alpha}$$

$$d_{23} = \frac{1}{|A_3|} P \sigma_{(\alpha-1)\alpha} |A_{3.1}|$$

$$d_{33} = \frac{1}{|A_3|} |A_{3.2}| \quad ,$$

where $A_{3.1}$ is the upper-left 1 x 1 matrix in A_3; $|A_{3.1}| = \sigma^2_{\alpha-2}$

$A_{3.2}$ is the upper-left 2 x 2 matrix in A_3:

$$|A_{3.2}| = \begin{vmatrix} \sigma^2_{\alpha-2} & -P\sigma_{(\alpha-2)(\alpha-1)} \\ -P\sigma_{(\alpha-2)(\alpha-1)} & \sigma^2_{\alpha-1} \end{vmatrix}$$

Hence, $a_{2w} = P^2 Q \, \sigma_{(\alpha-2)(\alpha-1)} \sigma_{(\alpha-1)\alpha} \dfrac{\sigma^2_{\alpha}}{|A_3|} = \dfrac{P^2 Q \rho_{(\alpha-2)(\alpha-1)} \rho_{(\alpha-1)\alpha}}{\Delta_3}$

$$\cdot \frac{\sigma_{\alpha-2}\sigma^2_{\alpha-1}\sigma_{\alpha}\sigma^2_{\alpha}}{\sigma^2_{\alpha-2}\sigma^2_{\alpha-1}\sigma^2_{\alpha}} = P^2 Q \frac{\sigma_{\alpha}}{\sigma_{\alpha-2}} \quad \frac{\rho_{(\alpha-2)(\alpha-1)}\rho_{(\alpha-1)\alpha}}{\Delta_3}$$

$$a_{1w} = PQ \, \sigma_{(\alpha-1)\alpha} \, \sigma^2_2 \, \frac{|A_{3.1}|}{|A_3|} = PQ \frac{\sigma_{\alpha}}{\sigma_{\alpha-1}} \quad \frac{\rho_{(\alpha-1)\alpha}}{\Delta_3} \qquad (5.4)$$

$$a_{ow} = Q\sigma^2_{\alpha} \quad \frac{|A_{3.2}|}{|A_3|} = Q \frac{\Delta_{3.2}}{\Delta_3} \quad ,$$

where Δ_3, $\Delta_{3.2}$ are defined as the determinants of the corresponding correlation matrices to A_3, A_{32} .

$$\Delta_3 = \begin{vmatrix} 1 & -P\rho_{(\alpha-2)(\alpha-1)} & 0 \\ -P\rho_{(\alpha-2)(\alpha-1)} & 1 & -P\rho_{(\alpha-1)\alpha} \\ 0 & -P\rho_{(\alpha-1)\alpha} & 1 \end{vmatrix}$$

$$\Delta_{3.2} = \begin{vmatrix} 1 & -P\rho_{(\alpha-2)(\alpha-1)} \\ -P\rho_{(\alpha-2)(\alpha-1)} & 1 \end{vmatrix}$$

$$\Delta_{3.1} = 1 \quad .$$

Hence the optimum estimator ${}_5\bar{Y}_{\alpha_2 w}$ and its variance are

$${}_5\bar{Y}_{\alpha_2 w} = \sum_{i=0}^{2} a_{iw}\, \bar{Y}_{(\alpha-i)1} + \sum_{i=0}^{2} b_{iw}\, \bar{Y}_{(\alpha-i)2} \tag{5.5}$$

$$V({}_5\bar{Y}_{\alpha_2 w}) = \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) + Q^2\,\frac{\Delta^2_{3.2}}{\Delta^2_3}\,\frac{N}{N-1}\,\frac{\sigma^2_\alpha}{nPQ} + P^2Q^2\,\frac{\sigma^2_\alpha}{\sigma^2_{\alpha-1}}\,\frac{\rho^2_{(\alpha-1)\alpha}}{\Delta^2_3}\,\frac{N}{N-1}\,\frac{\sigma^2_{\alpha-1}}{nPQ}$$

$$+ P^4Q^2\,\frac{\sigma^2_\alpha}{\sigma^2_{\alpha-2}}\,\frac{\rho^2_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha}}{\Delta^2_3}\,\frac{N}{N-1}\,\frac{\sigma^2_{\alpha-2}}{nPQ} - 2Q^2\,\frac{\Delta_{3.2}}{\Delta_3}\,\frac{N}{N-1}\,\frac{\sigma^2_\alpha}{nPQ}$$

$$- 2P^2Q^2\,\frac{\Delta_{3.2}}{\Delta^2_3}\,\frac{\sigma_\alpha\,\rho_{\alpha(\alpha-1)}}{\sigma_{\alpha-1}}\,\frac{N}{N-1}\,\frac{\sigma_{\alpha(\alpha-1)}}{nPQ}$$

$$- 2P^4Q^2\,\frac{\sigma^2_\alpha}{\sigma_{\alpha-1}\sigma_{\alpha-2}}\,\frac{\rho_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha}}{\Delta^2_3}\,\frac{N}{N-1}\,\frac{\sigma_{(\alpha-1)(\alpha-2)}}{nPQ}$$

$$= \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) + \frac{N}{N-1}\,\frac{\sigma^2_\alpha}{nPQ}\,\frac{Q^2}{\Delta^2_3}\,(\Delta^2_{3.2} + P^2\rho^2_{(\alpha-1)\alpha} +$$

$$+ P^4Q^2\rho^2_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha} - 2\Delta_{3.2}\Delta_3 - 2P^2\Delta_{3.2}\rho^2_{\alpha(\alpha-1)} -$$

$$- 2P^2\Delta_{3.2}\rho^2_{\alpha(\alpha-1)} - 2P^4\rho_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha}) \quad .$$

Since $\Delta_3 = \Delta_{3.2} - P^2\rho^2_{\alpha(\alpha-1)}$, we have

$$V({}_5\bar{\bar{Y}}_{\alpha_2 w}) = \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma^2_\alpha}{nPQ}\frac{Q^2}{\Delta^2_{3.2}}(\Delta^2_{3.2} - P^2\rho^2_{(\alpha-1)\alpha} + P^4\rho^2_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha}).$$

Letting $${}_5\phi_2 = \frac{1}{\Delta^2_3}(\Delta^2_{3.2} - P^2\rho^2_{(\alpha-1)\alpha} + P^4\rho^2_{(\alpha-2)(\alpha-1)}\rho^2_{(\alpha-1)\alpha}) \quad ,$$

we have

$$V({}_5\bar{\bar{Y}}_{\alpha_2 w}) = \frac{N}{N-1}\frac{\sigma^2_\alpha}{nPQ}(Q - Q^2\,{}_5\phi_2) - \frac{\sigma^2_\alpha}{N-1} \qquad (5.6)$$

(i = 3)

Similarly, the optimum estimator ${}_5\bar{\bar{Y}}_{\alpha_3 w}$ and its variance are as follows:

$${}_5\bar{\bar{Y}}_{\alpha_3 w} = \sum_{i=0}^{3} a_{iw}\bar{Y}_{(\alpha-i)1} + \sum_{i=0}^{3} b_{iw}\bar{Y}_{(\alpha-i)2}$$

where

$$a_{3w} = QP^3\sigma_{(\alpha-3)(\alpha-2)}\sigma_{(\alpha-2)(\alpha-1)}\sigma_{(\alpha-1)\alpha}\sigma^2_\alpha\frac{|A_{4.0}|}{|A_4|} =$$

$$= QP^3\rho_{(\alpha-3)(\alpha-2)}\rho_{(\alpha-2)(\alpha-1)}\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-3}}\ \frac{1}{\Delta_4}$$

$$a_{2w} = QP^2\sigma_{(\alpha-2)(\alpha-1)}\sigma_{(\alpha-1)\alpha}\sigma^2_\alpha\frac{|A_{4.1}|}{|A_4|} = QP^2\rho_{(\alpha-2)(\alpha-1)}\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-2}}\frac{\Delta_{4.1}}{\Delta_4}$$

$$a_{1w} = QP\,\sigma_{(\alpha-1)\alpha}\sigma^2_\alpha\frac{|A_{4.2}|}{|A_4|} = QP\,\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-1}}\frac{\Delta_{4.2}}{\Delta_4}$$

$$a_{ow} = Q\,\sigma^2_\alpha\frac{|A_{4.3}|}{|A_4|} = Q\frac{\Delta_{4.3}}{\Delta_4} \quad , \qquad (5.7)$$

where

$$|A_4| = \begin{vmatrix} \sigma^2_{\alpha-3} & -P\sigma_{(\alpha-3)(\alpha-2)} & 0 & 0 \\ -P\sigma_{(\alpha-3)(\alpha-2)} & \sigma^2_{\alpha-2} & -P\sigma_{(\alpha-2)(\alpha-1)} & 0 \\ 0 & -P\sigma_{(\alpha-2)(\alpha-1)} & \sigma^2_{\alpha-1} & -P\sigma_{\alpha(\alpha-1)} \\ 0 & 0 & -P\sigma_{\alpha(\alpha-1)} & \sigma^2_{\alpha} \end{vmatrix},$$

$$|A_{4.3}| = \begin{vmatrix} \sigma^2_{\alpha-3} & -P\sigma_{(\alpha-3)(\alpha-2)} & 0 \\ -P\sigma_{(\alpha-3)(\alpha-2)} & \sigma^2_{\alpha-2} & -P\sigma_{(\alpha-2)(\alpha-1)} \\ 0 & -P\sigma_{(\alpha-2)(\alpha-1)} & \sigma^2_{\alpha-1} \end{vmatrix}$$

$$|A_{4.2}| = \begin{vmatrix} \sigma^2_{\alpha-3} & -P\sigma_{(\alpha-3)(\alpha-2)} \\ -P\sigma_{(\alpha-3)(\alpha-2)} & \sigma^2_{\alpha-2} \end{vmatrix}$$

$$|A_{4.1}| = \sigma^2_{\alpha-3} \quad .$$

where Δ_4, $\Delta_{4.4-i}$ are defined as the determinants of the corresponding correlation matrices. Hence, the variance of the optimum estimator ${}_5\overline{Y}_{\alpha_3 W}$ is

$$V({}_5\overline{Y}_{\alpha_3 W}) = \frac{\sigma^2_\alpha}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma^2_\alpha}{nPQ}\frac{Q^2}{\Delta^2_4}\left(\Delta^2_{4.3} - P^2\rho^2_{\alpha,\alpha-1}\Delta^2_{4.2}\right.$$

$$\left. + P^4\rho_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\Delta^2_{4.1} - P^6\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\rho^2_{(\alpha-2)(\alpha-3)}\right) .$$

Let

$${}_5\phi_3 = \frac{1}{\Delta^2_4}\left(\Delta^2_{4.3} - P^2\rho^2_{\alpha(\alpha-1)}\Delta^2_{4.2} + P^4\rho^2_{(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\Delta^2_{4.1}\right.$$

$$\left. - P^6\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\rho^2_{(\alpha-2)(\alpha-3)}\right)$$

then

$$V({}_5\bar{Y}_{\alpha_3 w}) = \frac{N}{N-1}\frac{\sigma_\alpha^2}{nPQ}(Q - Q^2\ {}_5\phi_3) - \frac{\sigma_\alpha^2}{N-1} \quad . \tag{5.8}$$

(i - 1)

Let us consider a general case where the previous number of occasions from the αth occasion is $i - 1$. Then the variance and covariance matrix for the optimum solution of a_i is the matrix on the following page. The last column in A^{-1}, $\underline{d}_i$, is the following:

$$d_{1i} = P^{i-1}\ \sigma_{(\alpha-i+1)(\alpha-i+2)}\sigma_{(\alpha-i+2)(\alpha-i+3)}\cdots\sigma_{(\alpha-1)\alpha}\ \frac{1}{|A_i|}$$

$$d_{2i} = P^{i-2}\sigma_{(\alpha-i+2)(\alpha-i+3)}\sigma_{(\alpha-i+3)(\alpha-i+4)}\cdots\sigma_{(\alpha-1)\alpha}\ \frac{|A_{i.1}|}{|A_i|}$$

$$d_{3i} = P^{i-3}\sigma_{(\alpha-i+3)(\alpha-i+4)}\sigma_{(\alpha-i+4)(\alpha-i+5)}\cdots\sigma_{(\alpha-1)\alpha}\ \frac{|A_{i.2}|}{|A_i|}$$

$$\vdots$$

$$d_{i-j.i} = P^{j}\ \sigma_{(\alpha-j)(\alpha-j+1)}\sigma_{(\alpha-j+1)(\alpha-j+2)}\cdots\sigma_{(\alpha-1)\alpha}\ \frac{|A_i\,i-j-1|}{|A_i|}$$

$$\vdots \qquad \vdots$$

$$d_{i-2i} = P^2\sigma_{(\alpha-2)(\alpha-1)}\sigma_{(\alpha-1)\alpha}\ \frac{|A_{i(i-3)}|}{|A_i|}$$

$$d_{i-1i} = P\sigma_{(\alpha-1)\alpha}\ \frac{|A_{i\cdot i-2}|}{|A_i|}$$

$$d_{ii} = \frac{|A_{i\cdot i-1}|}{|A_i|}$$

where A_{ij} is the upper-left $j \times j$ matrix in A_i .

$$A_i = \begin{bmatrix}
\sigma^2_{\alpha-i+1} & -P\sigma_{(\alpha-i+1)(\alpha-i+2)} & 0 & & 0 & 0 & 0 & 0 \\
-P\sigma_{(\alpha-i+1)(\alpha-i+2)} & \sigma^2_{\alpha-i+2} & -P\sigma_{(\alpha-i+2)(\alpha-i+3)} & \cdots & 0 & 0 & 0 & 0 \\
0 & -P\sigma_{(\alpha-i+2)(\alpha-i+3)} & \sigma^2_{\alpha-i+3} & \cdots & 0 & 0 & 0 & 0 \\
 & \vdots & & & & \vdots & & \\
0 & 0 & & \cdots & \sigma^2_{\alpha-3} & -P\sigma_{(\alpha-1)(\alpha-2)} & 0 & 0 \\
0 & 0 & & \cdots & -P\sigma_{(\alpha-3)(\alpha-2)} & \sigma^2_{\alpha-2} & -P\sigma_{(\alpha-2)(\alpha-1)} & 0 \\
0 & 0 & & \cdots & 0 & -Pp_{(\alpha-2)(\alpha-1)} & \sigma^2_{\alpha-1} & -P\sigma_{(\alpha-1)\alpha} \\
0 & 0 & & \cdots & 0 & 0 & -P\sigma_{(\alpha-1)\alpha} & \sigma^2_{\alpha}
\end{bmatrix}$$

Since the solution vector $\underline{a} = Q\sigma_\alpha^2 \underline{d}_i$, we have

$$a_{i-1w} = Q\sigma_\alpha^2 d_{1i} = QP^{i-1}\sigma_{(2-i+1)(\alpha-i+2)}\sigma_{(\alpha-i+2)(\alpha-i+3)}\cdots\sigma_{(\alpha-2)(\alpha-1)}\sigma_\alpha^2 \frac{1}{|A_i|}$$

$$= QP^{i-1}\rho_{(\alpha-i+1)(\alpha-i+2)}\rho_{(\alpha-i+2)(\alpha-i+3)}\cdots\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-i+1}} \quad \frac{1}{\Delta_i}$$

$$a_{i-2w} = Q\sigma_\alpha^2 d_{2i} = QP^{i-2}\sigma_{(\alpha-i+2)(\alpha-i+3)}\cdots\sigma_{(\alpha-1)\alpha}\sigma_\alpha^2 \frac{|A_{i1}|}{|A_i|}$$

$$= QP^{i-2}\rho_{(\alpha-i+2)(\alpha-i+3)}\cdots\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-i+2}} \quad \frac{\Delta_{i\cdot 1}}{\Delta_i}$$

$$a_{i-3w} = Q\sigma_\alpha^2 d_{3i} = QP^{i-3}\sigma_{(\alpha-i+3)(\alpha-i+4)}\sigma_{(\alpha-i+4)(\alpha-i+5)}\cdots\sigma_{(\alpha-1)\alpha}\sigma_\alpha^2 \frac{|A_{i\cdot 2}|}{|A_i|}$$

$$= QP^{i-3}\rho_{(\alpha-i+3)(\alpha-i+4)}\cdots\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-i+3}} \quad \frac{\Delta_{i.2}}{\Delta_i}$$

$$\cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot$$

$$a_{jw} = Q\sigma_\alpha^2 d_{(i-j)i} = Qp^j\sigma_{(\alpha-j)(\alpha-j+1)}\sigma_{(\alpha-j+1)(\alpha-j+2)}\cdots\sigma_{(\alpha-1)\alpha}\sigma_\alpha^2 \frac{|A_{i.i-j-1}|}{|A_i|}$$

$$\cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot$$

$$a_{2w} = QP^2\sigma_{(\alpha-2)(\alpha-1)}\sigma_{(\alpha-1)\alpha} \quad \sigma_\alpha^2 \quad \frac{|A_{i\cdot i-3}|}{|A_i|} = QP^2\rho_{(\alpha-2)(\alpha-1)}\rho_{(\alpha-1)\alpha}$$

$$\frac{\sigma_\alpha}{\sigma_{\alpha-2}} \quad \frac{\Delta_{i\cdot i-3}}{\Delta_i}$$

$$a_{1w} = QP\ \sigma_{(\alpha-1)\alpha}\sigma_\alpha^2 \frac{|A_{i(i-2)}|}{|A_i|} = QP\rho_{(\alpha-1)\alpha}\frac{\sigma_\alpha}{\sigma_{\alpha-1}} \frac{\Delta_{i(1-2)}}{\Delta_i} \tag{5.9}$$

$$a_{ow} = Q\sigma_\alpha^2 \quad \frac{|A_{i\cdot i-1}|}{|A_i|} = Q\ \frac{\Delta_{i\cdot i-1}}{\Delta_i}$$

where $\Delta_i, \Delta_{i\cdot j}$ are the determinants of the corresponding correlation matrics to $A_i, A_{i\cdot j}$. For example

$$
\Delta_i = \begin{bmatrix}
1 & -P\rho_{(\alpha-i+1)(\alpha-i+2)} & 0 & 0 & 0 & 0 & 0 \\
-P\rho_{(\alpha-i+1)(\alpha-i+2)} & 1 & -P\rho_{(\alpha-i+2)(\alpha-i+3)} & & 0 & 0 & 0 \\
0 & -P\rho_{(\alpha-i+2)(\alpha-i+3)} & 1 & -P\rho_{(\alpha-i+3)(\alpha-i+4)} & 0 & 0 & 0 \\
\vdots & & \vdots & & & \ldots & \\
0 & 0 & \cdots & & -P\rho_{(\alpha-3)(\alpha-2)} & 0 & 0 \\
0 & 0 & \cdots & & 1 & -P\rho_{(\alpha-2)(\alpha-1)} & 0 \\
0 & 0 & \cdots & & -P\rho_{(\alpha-2)(\alpha-1)} & 1 & -P\rho_{(\alpha-1)\alpha} \\
0 & 0 & \cdots & & 0 & -P\rho_{(\alpha-1)\alpha} & 1
\end{bmatrix}
$$

Hence, we have

$$V({}_5\overline{Y}_{\alpha_{i-1}w}) = \frac{\sigma_\alpha^2}{Pn}\left(\frac{N-Pn}{N-1}\right) + \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{Q^2}{\Delta_i^2}\,(\Delta^2_{i\cdot i-1} + P^2\rho^2_{\alpha(\alpha-1)}\Delta^2_{i\cdot i-2}$$

$$+ P^4\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\Delta^2_{i(i-3)} + P^6\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\rho^2_{(\alpha-2)(\alpha-3)}$$

$$\cdot\,\Delta^2_{i\cdot i-4} + \cdots + P^{2j}\rho^2_{\alpha(\alpha-1)}\cdots\rho^2_{(\alpha-j+1)(\alpha-j)}\Delta^2_{i\cdot i-j-1}$$

$$\cdots + P^{2(i-2)}\rho^2_{\alpha(\alpha-1)}\cdots\,\rho^2_{(\alpha-i+3)(\alpha-i+2)}\,\Delta^2_{i\cdot 1}$$

$$+ P^{2(i-1)}\rho^2_{\alpha(\alpha-1)}\cdots\rho^2_{(\alpha-i+2)(\alpha-i+1)} - 2\Delta_i\Delta_{i\cdot i-1}$$

$$- 2P^2\rho^2_{\alpha(\alpha-1)}\Delta_{i\cdot i-1}\Delta_{i\cdot i-2} - 2\,P^4\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\Delta_{i\cdot i-2}\Delta_{i\cdot i-3})$$

$$- 2P^6\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\rho^2_{(\alpha-2)(\alpha-3)}\Delta_{i\cdot i-3}\Delta_{i\cdot i-4} - \cdots$$

$$- 2P^{2j}\rho^2_{\alpha(\alpha-1)}\,\cdots\,\rho^2_{(\alpha-j+2)(\alpha-j+1)}\rho^2_{(\alpha-j+1)(\alpha-j)}$$

$$\cdot\,\Delta_{i\cdot i-j-1}\Delta_{i\cdot i-j}\,\cdots\, - 2P^{2(i-2)}\rho^2_{\alpha(\alpha-1)}\,\cdots\,\rho^2_{(\alpha-i+3)(\alpha-i+2)}\Delta_{i\cdot 2}$$

$$- 2P^{2(i-1)}\rho^2_{\alpha(\alpha-1)}\,\cdots\,\rho^2_{(\alpha-i+2)(\alpha-i+1)}\Delta_{i\cdot 1}\,.$$

Since

$$\Delta_i = \Delta_{i\cdot i-1} - P^2\rho^2_{\alpha(\alpha-1)}\Delta_{i\cdot i-2}$$

$$\Delta_{i\cdot i-2} = \Delta_{i\cdot i-3} - P^2\rho^2_{(\alpha-2)(\alpha-3)}\Delta_{i\cdot i-4}$$

$$\vdots$$

$$\Delta_{i\cdot j} = \Delta_{i\cdot j-1} - P^2\rho^2_{(\alpha-i+j)(\alpha-i+j-1)}\Delta_{i\cdot j-2}$$

$$\vdots$$

$\Delta_{i.2} = \Delta_{i.1} - P^2\rho^2_{(\alpha-i+2)(\alpha-i+1)}\Delta_{i.o}$, the variance can be simplified.

$$V({}_5\overline{Y}_{\alpha_{i-1}w}) = \frac{\sigma_\alpha^2}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{Q^2}{\Delta_i^2}\,(\Delta^2_{i.i-1} - P^2\rho^2_{\alpha(\alpha-1)}\Delta^2_{i.i-2}$$

$$+ P^{2(2)}\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\Delta^2_{i.i-3} - P^{2(3)}\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)}\rho^2_{(\alpha-2)(\alpha-3)}\Delta^2_{i.i-4}$$

$$+ \ldots (-1)^j P^{2(j)}\rho^2_{\alpha(\alpha-1)}\rho^2_{(\alpha-1)(\alpha-2)} \cdots \rho^2_{(\alpha-j+1)(\alpha-j)}\Delta^2_{i.i-j-1} \cdots$$

$$(-1)^{i-1}P^{2(i-1)}\rho^2_{\alpha(\alpha-1)} \cdots \rho^2_{(\alpha-i+2)(\alpha-i+1)})$$

$$= \frac{\sigma_\alpha^2}{Pn}\left(\frac{N-Pn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{Q^2}{\Delta_i^2}\,\left(\Sigma\,(-1)^j P^{2j}\rho^2_{\alpha(\alpha-1)} \cdots \rho^2_{(\alpha-j+1)(\alpha-j)}\,\Delta^2_{i.i-j-1}\right) .$$

Letting $${}_5\phi_{i-1} = \frac{1}{\Delta_i^2}\left(\Sigma\,(-1)^j P^{2j}\rho^2_{\alpha(\alpha-1)} \cdots \rho^2_{(\alpha-j+1)(\alpha-j)}\,\Delta^2_{i.i-j-1}\right) .$$

We have

$$V({}_5\overline{Y}_{\alpha_{i-1}w}) = \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,(Q - Q^2\ {}_5\phi_{i-1}) - \frac{\sigma_\alpha^2}{N-1} . \tag{5.10}$$

$(i = \alpha - 1)$

In order to obtain the variance of the optimum estimator of $\overline{\overline{Y}}_\alpha$, using α-1 previous occasions, let $i = \alpha - 1$ in $V({}_5\overline{Y}_{\alpha_i w})$.

$$V({}_5\overline{Y}_{\alpha w}) = \frac{\sigma_\alpha^2}{Pn}\,\frac{N-Pn}{N-1} - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{Q^2}{\Delta_\alpha^2}\,\left(\sum_{j=0}^{\alpha-1}(-1)^j P^{2j}\rho^2_{\alpha(\alpha-1)}\right.$$

$$\left.\cdots\ \rho^2_{(\alpha-j+1)(\alpha-j)}\Delta^2_{\alpha-j-1}\right) .$$

Letting $${}_5\phi_{\alpha-1} = \frac{1}{\Delta_\alpha^2}\left(\sum_{j=0}^{\alpha-1}(-1)^j P^{2j}\rho^2_{\alpha(\alpha-1)} \cdots \rho^2_{(\alpha-j+1)(\alpha-j)}\Delta^2_{\alpha-j+1}\right) ,$$

we have

$$V({}_5\overline{Y}_{\alpha.w}) = \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,(Q - Q^2\ {}_5\phi_{\alpha-1}) - \frac{Q_\alpha^2}{N-1} . \tag{5.11}$$

<u>Relationship of percent loss of precision to number of previous occasions included in estimator</u>: Since the estimator ${}_5\bar{Y}_{\alpha_i}$ makes use of the information from i previous occasions, the sequence of variances of ${}_5\bar{Y}_{\alpha_i}$ is asymptotically decreasing as i increases. Hence, it is desired to determine the relationship of the loss of precision to the number of previous occasions included in the estimator.

Under the assumption $\rho_{(\alpha-i)(\alpha-i+1)} = \rho$, Table 5.1 provides percent loss of precision of an estimator ${}_5\bar{Y}_{\alpha_i}$ when i previous occasions are used in estimating $\bar{\bar{Y}}_\alpha$. The percent loss of precision is defined as

$$I_i = \frac{V({}_5\bar{Y}_{\alpha_i}) - V({}_5\bar{Y}_{\alpha.})}{V({}_5\bar{Y}_\alpha)} \cdot 100 \doteq \frac{s_i - s_{\alpha-1}}{s_{\alpha-1}} ,$$

where

$$s_i = Q - Q^2 \phi_i \quad \text{in (5.10)} .$$

Assuming $\alpha = 8$, loss of information of more than 5% arises only when $Q = .5$, $i = 1, 2$ and when $Q > .6$, $i = 1$. Therefore it is concluded that the estimator using one previous occasion has almost the same efficiency as the estimator using seven previous occasions over all Q larger than .6. When $Q = .5$, two previous occasions is preferable. Table 5.2 provides the optimum weights of the estimator using one or two previous occasions in estimating $\bar{\bar{Y}}_\alpha$. The values of a_{ow} attached to the sample mean $\bar{Y}_{\alpha 1}$ in (5.2). (5.5) are relatively higher than the values of a_{iw} for the previous occasions, and a_{1w} a_{2w} are rapidly decreasing. Since a_{ow} is directly related to ρ and Q, the sample mean $\bar{Y}_{\alpha 1}$ dominates the estimator

Table 5.1. Percent loss of precision of ${}_5\bar{\bar{Y}}_{\alpha_i}$

Q = .5

ρ \ I_i	I_1	I_2	I_3	I_4	I_5	I_6
.5	.55	.03	.00	.00	.00	.00
.6	1.37	.15	.01	.00	.00	.00
.7	3.25	.54	.09	.01	.00	.00
.8	7.93	1.95	.48	.12	.02	.00
.9	22.78	8.56	3.26	1.22	.43	.12
			Q = .6			
.5	.30	.01	.00	.00	.00	.00
.6	.70	.04	.00	.00	.00	.00
.7	1.52	.14	.01	.00	.00	.00
.8	3.15	.41	.05	.00	.00	.00
.9	6.53	1.17	.21	.03	.00	.00
			Q = .7			
.5	.13	.00	.00	.00	.00	.00
.6	.30	.01	.00	.00	.00	.00
.7	.61	.02	.00	.00	.00	.00
.8	1.17	.07	.00	.00	.00	.00
.9	2.14	.18	.01	.00	.00	.00
			Q = .8			
.5	.04	.00	.00	.00	.00	.00
.6	.09	.00	.00	.00	.00	.00
.7	.18	.00	.00	.00	.00	.00
.8	.32	.00	.00	.00	.00	.00
.9	.55	.01	.00	.00	.00	.00
			Q = .9			
.5	.00	.00	.00	.00	.00	.00
.6	.00	.00	.00	.00	.00	.00
.7	.00	.00	.00	.00	.00	.00
.8	.00	.00	.00	.00	.00	.00
.9	.00	.00	.00	.00	.00	.00

Table 5.2. Optimum values of a_i in $_5^{\bar{Y}}\alpha_i$

$_5^{\bar{Y}}\alpha_1 w$

Q	ρ \ a_i	a_{ow}	a_{1w}
Q = .5	.5	.533	.133
	.6	.549	.164
	.7	.569	.199
	.8	.595	.238
	.9	.626	.282
Q = .6	.5	.625	.125
	.6	.636	.152
	.7	.651	.182
	.8	.668	.213
	.9	.689	.248
Q = .7	.5	.716	.107
	.6	.723	.130
	.7	.732	.153
	.8	.742	.178
	.9	.755	.203
Q = .8	.5	.808	.080
	.6	.811	.097
	.7	.815	.114
	.8	.821	.131
	.9	.826	.148
Q = .9	.5	.902	.045
	.6	.903	.054
	.7	.904	.063
	.8	.905	.072
	.9	.907	.081

$_5^{\bar{Y}}\alpha_2 w$

Q	ρ \ a_i	a_{ow}	a_{1w}	a_{2w}
Q = .5	.5	.537	.142	.035
	.6	.554	.182	.054
	.7	.581	.231	.081
	.8	.617	.294	.117
	.9	.670	.378	.170
Q = .6	.5	.626	.130	.026
	.6	.639	.162	.039
	.7	.655	.199	.055
	.8	.677	.241	.077
	.9	.704	.291	.104
Q = .7	.5	.716	.109	.016
	.6	.724	.134	.024
	.7	.733	.161	.033
	.8	.745	.189	.045
	.9	.759	.221	.059
Q = .8	.5	.808	.081	.008
	.6	.811	.098	.011
	.7	.816	.116	.016
	.8	.821	.134	.021
	.9	.827	.153	.027
Q = .9	.5	.902	.045	.002
	.6	.903	.054	.003
	.7	.904	.063	.004
	.8	.905	.072	.005
	.9	.907	.082	.007

${}_5\bar{\bar{Y}}_{\alpha_1}$ when ρ and Q are high. Next, the percent gain of ${}_5\bar{\bar{Y}}_{\alpha_1}$ and ${}_5\bar{\bar{Y}}_{\alpha_2}$ over $\bar{\bar{Y}}_{\alpha}$ is presented in Table 5.3.

Table 5.3. Percent gain of ${}_5\bar{\bar{Y}}_{\alpha_1}$, ${}_5\bar{\bar{Y}}_{\alpha_2}$ over $\bar{\bar{Y}}_{\alpha}$

${}_5\bar{\bar{Y}}_{\alpha_1}$ over $\bar{\bar{Y}}_{\alpha}$

ρ \ Q	$Q=.5$	$Q=.6$	$Q=.7$	$Q=.8$	$Q=.9$
.5	7.14	6.66	5.67	4.21	2.30
.6	10.97	10.09	8.47	6.20	3.36
.7	16.22	14.62	12.06	8.69	4.63
.8	23.52	20.64	16.63	11.74	6.15
.9	34.03	28.75	22.47	15.46	7.93

${}_5\bar{\bar{Y}}_{\alpha_2}$ over $\bar{\bar{Y}}_{\alpha}$

ρ \ Q					
.5	7.69	6.97	5.81	4.25	2.31
.6	12.32	10.82	8.79	6.30	3.37
.7	19.36	16.20	12.72	8.88	4.66
.8	30.76	23.94	17.91	12.09	6.19
.9	51.59	35.57	24.86	16.08	8.00

The percent gain of ${}_5\bar{\bar{Y}}_{\alpha_1}$ is very high when Q is near .5 and it increases as ρ increases and as Q approaches .5. This can be understood because when $Q = .5$, ${}_5\bar{\bar{Y}}_{\alpha_1}$ is equivalent to the general linear estimator, and ${}_5\bar{\bar{Y}}_{\alpha_2}$ is a slight improvement over ${}_5\bar{\bar{Y}}_{\alpha_1}$.

The efficiency of the general linear estimator will be demonstrated by an example.

When we assume $\alpha = 3$ and $Q = \frac{2}{3}$, the optimum estimator ${}_1\bar{\bar{Y}}_{3w}$ is

$${}_1\bar{\bar{Y}}_{3w} = \bar{\bar{Y}}_{33} + \underline{a}'\bar{\underline{Y}}$$

where

$$\underline{a} = \begin{bmatrix} a_{3w} \\ b_{3w} \\ a_{2w} \\ b_{2w} \\ a_{1w} \\ b_{1w} \end{bmatrix} = \frac{1}{27-24\rho^2+4\rho^2} \begin{bmatrix} 9 - 4\rho^2 \\ 9 - 10\rho^2 + 2\rho^4 \\ 3\rho \\ 3\rho - 2\rho^3 \\ \rho^2 \\ \rho^2 \end{bmatrix}$$

$$\underline{\bar{Y}} = \begin{bmatrix} \bar{Y}_{31} - \bar{Y}_{33} \\ \bar{Y}_{32} - \bar{Y}_{33} \\ \bar{Y}_{21} - \bar{Y}_{23} \\ \bar{Y}_{22} - \bar{Y}_{23} \\ \bar{Y}_{11} - \bar{Y}_{13} \\ \bar{Y}_{12} - \bar{Y}_{13} \end{bmatrix}$$

and the variance of ${}_1\bar{Y}_{3w}$ is obtained as

$$V({}_1\bar{Y}_{3w}) = 3\,\frac{N}{N-1}\,\frac{\sigma_3^2}{n}\left(1 - \frac{n}{3N} - \frac{18-14\rho^2+2\rho^4}{27-24\rho^2+4\rho^4}\right) .$$

Thus

$$V({}_1\bar{Y}_{3w}) = \frac{3N}{N-1}\,\frac{\sigma_3^2}{n}\,(.312) \quad \text{when} \quad \rho = .5$$

$$= \frac{3N}{N-1}\,\frac{\sigma_3^2}{n}\,(.218) \quad \text{when} \quad \rho = .9 .$$

From (5.6)

$$V({}_5\bar{Y}_{3w}) = \frac{3N}{N-1}\,\frac{\sigma_3^2}{n}\,(.314) \quad \text{when} \quad \rho = .5$$

$$= \frac{3N}{N-1}\,\frac{\sigma_3^2}{n}\,(.261) \quad \text{when} \quad \rho = .9 .$$

The relative efficiency of ${}_5\bar{\bar{Y}}_{3w}$ to ${}_1\bar{\bar{Y}}_{3w}$ is .994 when $\rho = .5$ and .839 when $\rho = .9$.

5.3. A Modified Composite Estimator and Its Properties

A modified composite estimator recursively defined is

$${}_2\bar{\bar{Y}}_{\alpha.} = a_\alpha({}_2\bar{\bar{Y}}_{\alpha-1.} + \bar{\bar{Y}}_{\alpha 1} - \bar{\bar{Y}}_{(\alpha-1)\alpha}) + (1-a_\alpha)\bar{\bar{Y}}_{\alpha 2}$$

where, $\bar{\bar{Y}}_{(\alpha-1)\alpha}$ is the sample mean on the $\alpha-1^{th}$ occasion based on the matched portion for both the $\alpha-1^{th}$ and α^{th} occasions. In the formula, the term ${}_2\bar{\bar{Y}}_{\alpha-1.}$, an estimator of $\bar{\bar{\bar{Y}}}_{\alpha-1}$, is defined just as is ${}_2\bar{\bar{Y}}_{\alpha.}$.

The variance formula of ${}_2\bar{\bar{Y}}_{\alpha.}$ changes form rather drastically as the discarded fraction Q varies over the value of .1, .2, .3, .4 and a value larger than .5. In 5.3.1, the variance of ${}_2\bar{\bar{Y}}_{\alpha.}$ will be derived for $Q \geq \frac{1}{2}$. In 5.3.2., the variance of ${}_2\bar{\bar{Y}}_{\alpha_2}$ will be derived for $Q \leq \frac{1}{2}$.

5.3.1. Properties of ${}_2\bar{\bar{Y}}_{\alpha.}$ when $Q \geq \frac{1}{2}$

In order to find the variance of ${}_2\bar{\bar{Y}}_{\alpha.}$, let us start from a simple case, that is, $\alpha = 2$.

($\alpha = 2$)

$${}_2\bar{\bar{Y}}_2 = a_2({}_2\bar{\bar{Y}}_1 + \bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{12}) + (1 - a_2)\bar{\bar{Y}}_{22}$$

$$= \bar{\bar{Y}}_{22} - a_2(\bar{\bar{Y}}_{21} - \bar{\bar{Y}}_{22}) + a_2 Q(\bar{\bar{Y}}_{11} - \bar{\bar{Y}}_{12}) .$$

This was already found in 4.2.2. Hence from (4.8) the optimum value of a_2 is

$$a_{2w} = \frac{P\,\sigma_2^2}{\sigma_2^2 + \sigma^2\sigma_1^2 - 2Q^2\sigma_{12}}$$

and the optimum estimator ${}_2\overline{\overline{Y}}_{2w}$ and its variance are, from (4.9)

$$ {}_2\overline{\overline{Y}}_{2w} = a_{2w}({}_2\overline{\overline{Y}}_1 + \overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{12}) + (1 - a_{2w})\overline{\overline{Y}}_{22} \tag{5.13} $$

$$ V({}_2\overline{\overline{Y}}_{2w}) = \frac{\sigma_2^2}{Qn}\,\frac{N - Qn}{N - 1} - \frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ}\,\frac{P^2\sigma_2^2}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}} . $$

Hence, if the number of previous occasions from the αth occasion is one, i.e., $i = 1$ then,

$$ V({}_2\overline{\overline{Y}}_{\alpha_1 w}) = \frac{\sigma_\alpha^2}{Qn}\,\frac{N-Qn}{N-1} - \frac{\sigma_\alpha^2}{nPQ}\,\frac{P^2\sigma_\alpha^2}{\sigma_\alpha^2 + Q^2\sigma_{\alpha-1}^2 - 2Q^2\sigma_{\alpha(\alpha-1)}} \tag{5.14} $$

$(\alpha = 3)$

$$ {}_2\overline{\overline{Y}}_3 = a_3({}_2\overline{\overline{Y}}_2 + \overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{23}) + (1-a_3)\overline{\overline{Y}}_{32} \quad \text{and} $$

substituting ${}_2\overline{\overline{Y}}_2$ given in (5.13)

$$ {}_2\overline{\overline{Y}}_3 = a_3\left[\overline{\overline{Y}}_{22} + a_2(\overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{22}) + a_2Q(\overline{\overline{Y}}_{11} - \overline{\overline{Y}}_{12}) + (\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{23})\right] + (1-a_3)\overline{\overline{Y}}_{32} $$

$$ = \overline{\overline{Y}}_{32} + a_3(\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}) + a_3(\overline{\overline{Y}}_{22} - \overline{\overline{Y}}_{23}) + a_3a_2(\overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{22}) $$

$$ + a_3a_2Q(\overline{\overline{Y}}_{11} - \overline{\overline{Y}}_{12}) $$

$$ V({}_2\overline{\overline{Y}}_3) = V(\overline{\overline{Y}}_{32}) + a_3^2\,V(\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}) + a_3^2\,V(\overline{\overline{Y}}_{22} - \overline{\overline{Y}}_{23}) + a_3^2a_2^2\,V(\overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{22}) $$

$$ + a_3^2a_2^2Q^2V(\overline{\overline{Y}}_{11} - \overline{\overline{Y}}_{12}) + 2a_3\text{cov}(\overline{\overline{Y}}_{32},\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}) + 2a_3\text{cov}(\overline{\overline{Y}}_{32},\overline{\overline{Y}}_{22} - \overline{\overline{Y}}_{23}) $$

$$ + 2a_3a_2\text{cov}\,(\overline{\overline{Y}}_{32}, \overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{22}) + 2a_3a_2Q\text{ cov }(\overline{\overline{Y}}_{32}, \overline{\overline{Y}}_{11} - \overline{\overline{Y}}_{12}) $$

$$ + 2a_3^2\text{ cov }(\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}, \overline{\overline{Y}}_{22} - \overline{\overline{Y}}_{23}) + 2a_3^2a_2\text{ cov }(\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}, \overline{\overline{Y}}_{21} - \overline{\overline{Y}}_{22}) $$

$$ + 2a_3^2a_2Q\text{ cov }(\overline{\overline{Y}}_{31} - \overline{\overline{Y}}_{32}, \overline{\overline{Y}}_{11} - \overline{\overline{Y}}_{12}) + 2a_2^2a_2\text{ cov }(\overline{\overline{Y}}_{22} - \overline{\overline{Y}}_{23},\overline{\overline{Y}}_{21}, - \overline{\overline{Y}}_{22}) $$

$$ + 2a_3^2a_2Q\text{ cov }(\overline{\overline{Y}}_{22}-\overline{\overline{Y}}_{23},\overline{\overline{Y}}_{11}-\overline{\overline{Y}}_{12}) + 2a_3^2a_2^2Q\text{ cov }(\overline{\overline{Y}}_{21}-\overline{\overline{Y}}_{22},\overline{\overline{Y}}_{11}-\overline{\overline{Y}}_{12}) . $$

Figure 5.2 may help in the derivation of variances and covariances, where sample means and their sizes are indicated. The variances and covariances in $V({}_2\bar{Y}_3)$ are as follows:

$$V(\bar{Y}_{31} - \bar{Y}_{32}) = \frac{N}{N-1} \frac{\sigma_3^2}{nPQ}$$

$$V(\bar{Y}_{22} - \bar{Y}_{23}) = V(\frac{Q-P}{Q} \bar{Y}_{20} + \frac{P}{Q} \bar{Y}_{23} - \bar{Y}_{23}) = \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} (Q-P) .$$

The proof is given in the Appendix 9.4.

$$\text{cov } (\bar{Y}_{32}, \bar{Y}_{31} - \bar{Y}_{32}) = -\frac{N}{N-1} \frac{\sigma_3^2}{Qn}$$

$$\text{cov } (\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{22} - \bar{Y}_{23}) = \text{cov } \bar{Y}_{31}\bar{Y}_{22} - \text{cov } \bar{Y}_{31}\bar{Y}_{23} = \frac{N}{N-1} \frac{\sigma_{23}}{nPQ} (P-Q)$$

The proof is also given in the Appendix 9.5.

$$\text{cov } (\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{21} - \bar{Y}_{22}) = -\frac{N}{N-1} \frac{\sigma_{23}}{Qn}$$

$$\text{cov } (\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{11} - \bar{Y}_{12}) = 0$$

$$\text{cov } (\bar{Y}_{22} - \bar{Y}_{23}, \bar{Y}_{21} - \bar{Y}_{22}) = 0$$

$$\text{cov } (\bar{Y}_{21} - \bar{Y}_{22}, \bar{Y}_{11} - \bar{Y}_{12}) = -\frac{N}{N-1} \frac{\sigma_{12}}{pn} .$$

Hence

$$V({}_2\bar{Y}_3) = \frac{\sigma_3^2}{Qn} (\frac{N-Qn}{N-1}) + a_3^2 \frac{N}{N-1} \frac{\sigma_3^2}{nPQ} + a_3^2 \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} (Q-P) + a_3^2 a_2^2 \frac{N}{N-1} \frac{\sigma_2^2}{nPQ}$$

$$+ a_3^2 a_2^2 Q^2 \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} - 2a_3 \frac{N}{N-1} \frac{\sigma_3^2}{Qn} + 2a_3^2 \frac{N}{N-1} \frac{\sigma_{23}}{nPQ} (P-Q)$$

$$- 2a_3^2 a_2 \frac{N}{N-1} \frac{\sigma_{23}}{Qn} - 2a_3^2 a_2^3 Q^2 \frac{N}{N-1} \frac{\sigma_{12}}{nPQ} .$$

<u>Optimum estimator and its variance ($\alpha = 3$)</u>: The equations for the optimum values of a_i are

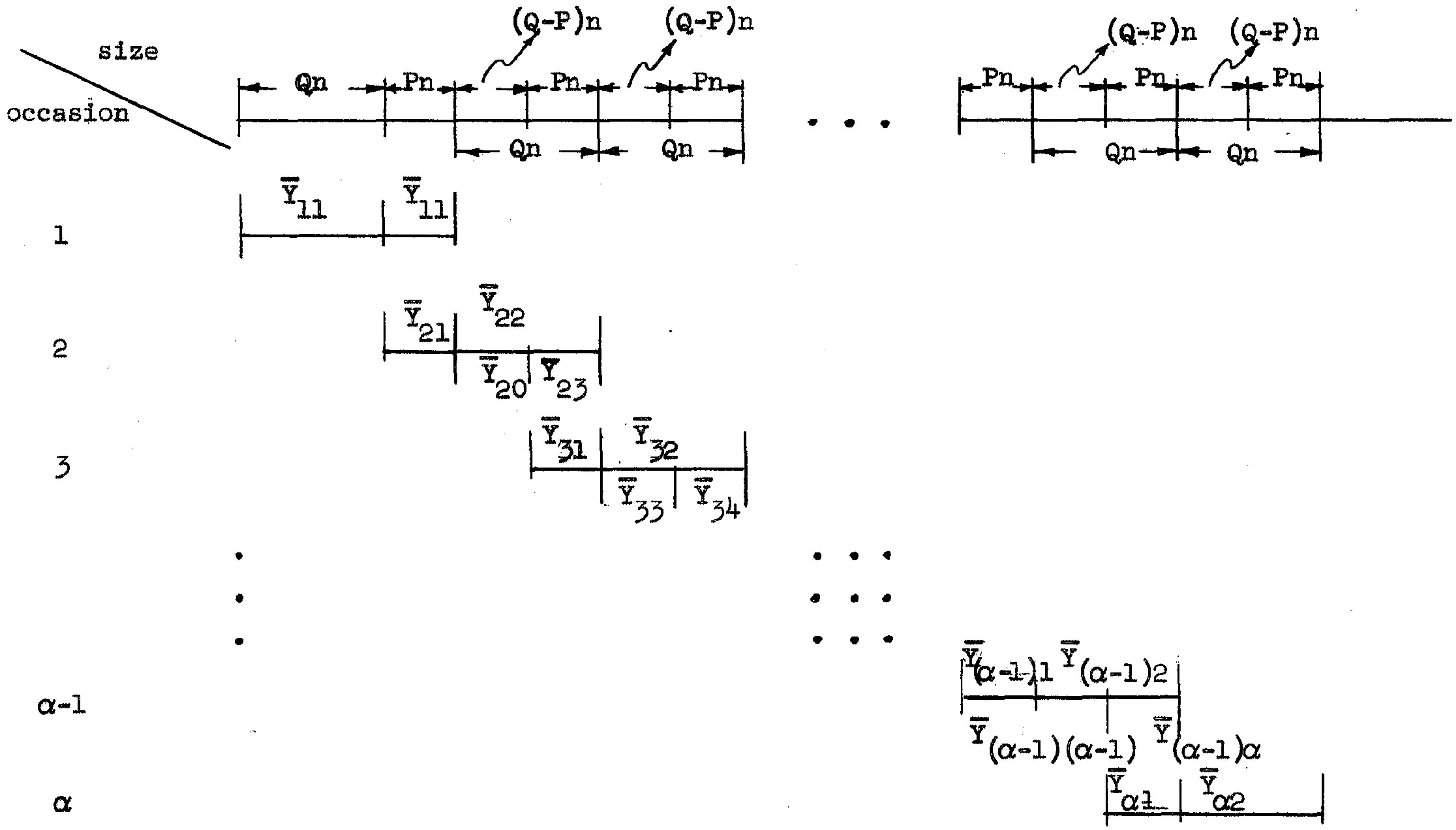


Figure 5.2. Formation of $_{2}\bar{Y}_{\alpha.}$ when $Q \geq \frac{1}{2}$

$$\frac{(N-1)}{2}\frac{nPQ}{N}\frac{\partial}{\partial a_2} V({}_2\bar{Y}_3) = a_2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}) - P\sigma_{23} = 0$$

$$\frac{(N-1)}{2}\frac{nPQ}{N}\frac{\partial}{\partial a_3} V({}_2\bar{Y}_3) = a_3\Big[\sigma_3^2 + \sigma_2^2(Q-P) + a_2^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12})$$

$$- 2\sigma_{23}(Q-P+a_2P)\Big] - P\sigma_3^2 = 0 \quad .$$

From these,

$$a_{2w} = \frac{P\sigma_{23}}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}} \tag{5.15}$$

$$a_{3w} = \frac{P\sigma_3^2}{\sigma_3^2 + \sigma_2^2(Q-P) - a_{2w}P\sigma_{23} - 2\sigma_{23}(Q-P)} \quad .$$

Substituting a_{2w} and a_{3w} in ${}_2\bar{Y}_3$ and $V({}_2\bar{Y}_3)$, we have the optimum estimator ${}_2\bar{Y}_{3w}$ and its variance:

$$V({}_2\bar{Y}_{3w}) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_3^2}{nPQ}Pa_{3w}$$

$$= \frac{N}{N-1}\frac{\sigma_3^2}{nPQ}(P-Pa_{3w}) - \frac{\sigma_3^2}{N-1} \quad . \tag{5.16}$$

The proof is given in the Appendix 9.6. Hence

$$V({}_2\bar{Y}_{\alpha_2 w}) = \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_\alpha^2}{nPQ}Pa_{\alpha w} \quad .$$

($\alpha=4$)

$${}_2\bar{Y}_4 = a_4({}_2\bar{Y}_3 + \bar{Y}_{41} - \bar{Y}_{34}) + (1 - a_4)\bar{Y}_{42}$$

or

$$_2\overline{Y}_4 = \overline{Y}_{42} + a_4(\overline{Y}_{41} - \overline{Y}_{42}) + a_4\ (\overline{Y}_{32} - \overline{Y}_{34}) + a_4a_3(\overline{Y}_{31} - \overline{Y}_{32})$$

$$+ a_4a_3(\overline{Y}_{22} - \overline{Y}_{23}) + a_4a_3a_2(\overline{Y}_{21} - \overline{Y}_{22}) + a_4a_3a_2Q(\overline{Y}_{11} - \overline{Y}_{12}).$$

The necessary computations of variances and covariances for the derivation of $V(_2\overline{Y}_4)$ were given in the previous case $\alpha = 3$.

$$V(_2\overline{Y}_4) = \frac{\sigma_4^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + a_4^2\ \frac{N}{N-1}\ \frac{\sigma_4^2}{nPQ} + a_4^2\ \frac{N}{N-1}\ \frac{\sigma_3^2}{nPQ}\ (Q-P) + a_4^2a_3^2\ \frac{N}{N-1}\ \frac{\sigma_3^2}{nPQ}$$

$$+ a_4^2a_3^2\ \frac{N}{N-1}\ \frac{\sigma_2^2}{nPQ}\ (Q-P) + a_4^2a_3^2a_2^2\ \frac{N}{N-1}\ \frac{\sigma_2^2}{nPQ} + a_4^2a_3^2a_2^2Q^2\ \frac{N}{N-1}\ \frac{\sigma_1^2}{nPQ}$$

$$- 2\ a_4\ \frac{N}{N-1}\ \frac{\sigma_4^2}{Qn} + 2\ a_4^2\ \frac{N}{N-1}\ \frac{\sigma_{34}}{nPQ}\ (P-Q) - 2a_4^2a_3\ \frac{N}{N-1}\ \frac{\sigma_{34}}{Qn}$$

$$+ 2\ a_4^2a_3^2\ \frac{N}{N-1}\ \frac{\sigma_{23}}{nPQ}\ (P-Q) - 2\ a_4^2a_3^2a_2\ \frac{N}{N-1}\ \frac{\sigma_{23}}{Qn}$$

$$- 2\ a_4^2a_3^2a_2^2Q^2\ \frac{N}{N-1}\ \frac{\sigma_{12}}{nPQ}\ .$$

<u>Optimum estimator and its variance ($\alpha = 4$)</u>: The optimum values of a_2, a_3 and a_4 are obtained as follows:

$$a_{2w} = \frac{P\sigma_{23}}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}}$$

$$a_{3w} = P\sigma_{34}/[\,\sigma_3^2 + \sigma_2^2(Q-P) + a_{2w}^2\sigma_2^2 + a_2^2Q^2\sigma_1^2 + 2\sigma_{23}(P-Q)$$

$$- 2a_{2w}P\sigma_{23} - 2a_2^2Q^2\sigma_{12}],\ \text{since}\ \sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12} = \frac{P\sigma_{23}}{a_{2w}}$$

$$a_{3w} = \frac{P\sigma_{34}}{\sigma_3^2 + \sigma_2^2(Q-P) - a_{2w}P\sigma_{23} - 2\sigma_{23}(Q-P)}$$

$$a_{4w} = P\sigma_4^2/[\sigma_4^2 + \sigma_3^2(Q-P) + a_3^2\sigma_3^2 + a_2^2\sigma_2^2(Q-P)$$

$$+ a_2^2a_2^2\sigma_2^2 + a_3^2a_2^2Q^2\sigma_1^2 + 2\sigma_{34}(P-Q) - 2a_3P\sigma_{34}$$

$$+ 2a_3^2\sigma_{23}(P-Q) - 2a_3^2a_2P\sigma_{23} - 2a_3^2a_2^2Q_2\sigma_{12}] \quad .$$

From a_{2w} and a_{3w}

$$a_{4w} = \frac{P\sigma_4^2}{\sigma_4^2 + \sigma_3^2(Q-P) + a_3P\sigma_{34} - 2\sigma_{34}(Q-P) - 2a_3P\sigma_{34}}$$

$$= \frac{P\sigma_4^2}{\sigma_4^2 + \sigma_3^2(Q-P) - a_{3w}P\sigma_{34} - 2\sigma_{34}(Q-P)} \quad . \qquad (5.17)$$

The variance of ${}_2\overline{Y}_{4w}$ becomes

$$V({}_2\overline{Y}_{4w}) = \frac{\sigma_4^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_4^2}{nPQ}Pa_{4w} \qquad (5.18)$$

$$= \frac{N}{N-1}\frac{\sigma_4^2}{nPQ}(P - Pa_{4w}) - \frac{\sigma_4^2}{N-1} \quad .$$

(α-5)

$${}_2\overline{Y}_5 = a_5({}_2\overline{Y}_4 + \overline{Y}_{51} - \overline{Y}_{45}) + (1 - a_5)\overline{Y}_{52}$$

$$= \overline{Y}_{52} + a_5(\overline{Y}_{51} - \overline{Y}_{52}) + a_5(\overline{Y}_{42} - \overline{Y}_{45}) + a_5a_4(\overline{Y}_{41} - \overline{Y}_{42})$$

$$+ a_5a_4(\overline{Y}_{32} - \overline{Y}_{34}) + a_5a_4a_3(\overline{Y}_{31} - \overline{Y}_{32}) + a_5a_4a_3(\overline{Y}_{22} - \overline{Y}_{23})$$

$$+ a_5a_4a_3a_2(\overline{Y}_{21} - \overline{Y}_{22}) + a_5a_4a_3a_2Q(\overline{Y}_{11} - \overline{Y}_{12}) \quad .$$

With the same argument as for ${}_2\overline{Y}_{4w}$, the optimum estimator and its variance are as follows:

$$a_{2w} = \frac{P\sigma_{23}}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}}$$

$$a_{3w} = \frac{P\sigma_{34}}{\sigma_3^2+\sigma_2^2(Q-P)-a_{2w}P\sigma_{23}-2\sigma_{23}(Q-P)}$$

$$a_{4w} = \frac{P\sigma_{45}}{\sigma_4^2 + \sigma_3^2(Q-P)-a_{3w}P\sigma_{34}-2\sigma_{34}(Q-P)} \tag{5.19}$$

$$a_{5w} = \frac{P\sigma_5^2}{\sigma_5^2+\sigma_4^2(Q-P)-a_{4w}P\sigma_{45}-2\sigma_{45}(Q-P)}$$

$$V({}_2\overline{Y}_{5w}) = \frac{\sigma_5^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\ \frac{\sigma_5^2}{nPQ}\ Pa_{5w} \tag{5.20}$$

$$= \frac{N}{N-1}\ \frac{\sigma_5^2}{nPQ}\ (P-Pa_{5w}) - \frac{\sigma_5^2}{N-1}$$

$(\alpha=\alpha)$

Finally, for any finite number α, an estimator of $\overline{\overline{Y}}_\alpha$ is

$${}_2\overline{Y}_{\alpha.} = a_\alpha({}_2\overline{Y}_{\alpha-1}+ \overline{Y}_{\alpha 1} - \overline{Y}_{(\alpha-1)\alpha}) + (1-a_\alpha)\overline{Y}_{\alpha 2}$$

$$= \overline{Y}_{\alpha 2} + a_\alpha(\overline{Y}_{\alpha 1} - \overline{Y}_{\alpha 2}) + a_\alpha(\overline{Y}_{(\alpha-1)2} - \overline{Y}_{(\alpha-1)\alpha}) + a_\alpha a_{\alpha-1}(\overline{Y}_{(\alpha-1)1}-\overline{Y}_{(\alpha-1)2})$$

$$+ a_\alpha a_{\alpha-1}(\overline{Y}_{(\alpha-1)2} - \overline{Y}_{(\alpha-2)(\alpha-1)}) + a_\alpha a_{\alpha-1}a_{\alpha-2}(\overline{Y}_{(\alpha-2)1} - \overline{Y}_{(\alpha-2)2})$$

$$\vdots \qquad \vdots$$

$$+ a_\alpha a_{\alpha-1} \cdots a_{\alpha-i+1}(\overline{Y}_{(\alpha-i)2} - \overline{Y}_{(\alpha-i)(\alpha-i+1)}) + a_\alpha a_{\alpha-1}\cdots a_{\alpha-i}$$

$$\cdot(\overline{Y}_{(\alpha-i)1} - \overline{Y}_{(\alpha-i)2})$$

$$\vdots \qquad \vdots$$

$$+ a_\alpha \cdots a_3 a_2 (\bar{Y}_{21} - \bar{Y}_{22}) + a_\alpha \cdots a_2 Q(\bar{Y}_{11} - \bar{Y}_{12})$$

$$= \bar{Y}_\alpha + a_\alpha(\bar{Y}_{\alpha 1} - \bar{Y}_{\alpha 2}) + \sum_{i=1}^{\alpha-2} a_\alpha a_{\alpha-i} \cdots a_{\alpha-i+1} (\bar{Y}_{(\alpha-i)2} - \bar{Y}_{(\alpha-i)(\alpha-i+1)})$$

$$+ \sum_{i=1}^{\alpha-2} a_\alpha a_{\alpha-1} \cdots a_{\alpha-i} (\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2})$$

$$+ a_\alpha \cdots a_2 Q(\bar{Y}_{11} - \bar{Y}_{12}) \quad .$$

The variance of ${}_2\bar{Y}_\alpha$ is given in Appendix 9.7. The optimum estimator and its variance are found as before. That is

$$a_{2w} = \frac{P\sigma_{23}}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}}$$

$$a_{3w} = \frac{P\sigma_{34}}{\sigma_3^2 + \sigma_2^2(Q-P) - 2\sigma_{23}(Q-P) - a_{2w} P\sigma_{23}}$$

$$a_{iw} = \frac{P\sigma_{i(i+1)}}{\sigma_i^2 + \sigma_{i-1}^2(Q-P) - 2\sigma_{i(i-1)}(Q-P) - a_{(i-1)w} P\sigma_{i(i-1)}} \qquad (5.21)$$

$$i = (4, 5, \ldots, \alpha-1)$$

$$a_{\alpha w} = \frac{P\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_{\alpha-1}^2(Q-P) - 2\sigma_{\alpha(\alpha-1)}(Q-P) - a_{(\alpha-1)w} P\sigma_{\alpha(\alpha-1)}}$$

and

$$V({}_2\bar{Y}_{\alpha.w}) = \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,Pa_{\alpha w} = \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,(P - Pa_{\alpha w}) - \frac{\sigma_\alpha^2}{N-1} \quad . (5.22)$$

<u>Relationship of percent loss of precision to number of previous occasions included in estimator:</u> For a managerable numerical analysis, the maximum α was chosen equal to 5. It was also assumed that

$\sigma^2_{\alpha-i} = \sigma^2$ and $\rho_{(\alpha-i)(2-i+1)} = \rho$ for every occasion. Table 5.4 provides percent loss of precision of an estimator ${}_2\overline{Y}_{\alpha_i}$ when i previous occasions are used in estimating $\overline{\overline{Y}}_\alpha$ and the optimum value of a_i in ${}_2\overline{Y}_{\alpha_2 w}$.

The percent loss is defined as before, i.e.,

$$I_i = \frac{V({}_2\overline{Y}_{\alpha_i w}) - V({}_2\overline{Y}_{\alpha_4 w})}{V({}_2\overline{Y}_{\alpha_4 w})} \quad 100 \doteq \frac{M_i - M_4}{M_4}$$

where $M_i = \mathbf{P} - \mathbf{P}a_{\alpha w}$ in (5.16), (5.18). From this numerical analysis, it may be seen that under most conditions the preferred number of previous occasions is two. In Table 5.4, the optimum weights of the estimator of $\overline{\overline{Y}}_\alpha$ using two previous occasions are given in the following form. The estimator ${}_2\overline{Y}_{\alpha_2}$ can be written as

$${}_2\overline{Y}_{\alpha_2} = a_\alpha({}_2\overline{Y}_{\alpha_1} + \overline{Y}_{\alpha_1} - \overline{Y}_{(\alpha-1)\alpha}) + (1-a_\alpha)\overline{Y}_{\alpha 2}$$

$$= \overline{Y}_{\alpha 2} + a_\alpha(\overline{Y}_{\alpha 1} - \overline{Y}_{\alpha 2}) + a_\alpha(\overline{Y}_{(\alpha-1)2} - \overline{Y}_{(\alpha-1)\alpha})$$

$$+ a_\alpha a_{\alpha-1}(\overline{Y}_{(\alpha-1)1} - \overline{Y}_{(\alpha-1)2}) + a_\alpha a_{\alpha-1}Q(\overline{Y}_{(\alpha-2)1} - \overline{Y}_{(\alpha-2)2}).$$

The coefficients can be denoted

$$a_{\alpha w}a_{(\alpha-1)w} = (a_{\alpha-1})_w \quad \text{and} \quad a_{\alpha w}a_{(\alpha-1)w}Q = (a_{\alpha-2})_w \quad .$$

From Table 5.4 it can be seen that $a_{\alpha w}$, the weight of the estimator $({}_2\overline{Y}_{\alpha_1} + \overline{Y}_{\alpha 1} - \overline{Y}_{(\alpha-1)\alpha})$, is increasing directly related to ρ and increasing as the matched fraction **P** approaches .5. The weights attached to the difference on occasion α-1 and α-2 are repidly decreasing.

Table 5.4. Percent loss of precision of ${}_2\bar{Y}_{\alpha_i}$ and optimum value of a_α when $\alpha = 3$

ρ \ I_i	I_1	I_2	I_3	ρ	$(a_{\alpha-2})_w$	$(a_{\alpha-1})_w$	$a_{\alpha w}$
Q = .5							
.5	6.73	.54	.03	.5	.066	.133	.533
.6	6.56	.71	.07	.6	.087	.174	.552
.7	6.64	1.07	.15	.7	.112	.225	.578
.8	9.63	2.26	.45	.8	.144	.289	.615
.9	21.89	7.44	2.06				
Q = .6							
.5	2.99	.12	.00	.5	.049	.083	.416
.6	2.80	.20	.01	.6	.069	.115	.445
.7	3.40	.38	.03	.7	.094	.157	.482
.8	6.22	1.11	.17	.8	.130	.217	.533
.9	17.32	4.97	1.19	.9	.184	.307	.607
Q = .7							
.5	1.02	.02	.00	.5	.032	.046	.306
.6	1.06	.04	.00	.6	.047	.067	.339
.7	1.56	.11	.00	.7	.069	.099	.382
.8	3.55	.44	.02	.8	.105	.150	.442
.9	12.08	2.74	.53	.9	.166	.237	.535
Q = .8							
.5	.25	.00	.00	.5	.016	.020	.202
.6	.29	.00	.00	.6	.025	.031	.231
.7	.54	.01	.00	.7	.041	.051	.272
.8	1.54	.10	.00	.8	.069	.086	.334
.9	6.65	1.03	.14	.9	.130	.162	.440
Q = .9							
.5	.02	.00	.00	.5	.004	.005	.100
.6	.03	.00	.00	.6	.007	.008	.119
.7	.08	.00	.00	.7	.013	.015	.148
.8	.31	.00	.00	.8	.027	.030	.197
.9	1.94	.13	.00	.9	.068	.075	.296

Next, the percent gain of ${}_2\overline{\overline{Y}}_{\alpha_1}$ and ${}_2\overline{\overline{Y}}_{\alpha_2}$ over $\overline{\overline{Y}}_{\alpha}$ is presented in Table 5.5.

Table 5.5. Percent gain of ${}_2\overline{\overline{Y}}_{\alpha_1}$, ${}_2\overline{\overline{Y}}_{\alpha_2}$ over $\overline{\overline{Y}}_{\alpha}$

${}_2\overline{\overline{Y}}_{\alpha_1}$ over $\overline{\overline{Y}}_{\alpha}$

ρ \ Q	.2	.3	.4	.5	.6	.7	.8	.9
.5	0.	0.	0.	0.	0.	0.	0.	0.
.6	3.33	4.46	5.21	5.55	5.45	4.88	3.80	2.19
.7	6.95	9.54	11.42	12.50	12.63	11.66	9.41	5.62
.8	10.90	15.36	18.94	21.42	22.50	21.72	18.46	11.73
.9	15.23	22.10	28.23	33.15	36.92	38.18	35.55	25.71

${}_2\overline{\overline{Y}}_{\alpha_2}$ over $\overline{\overline{Y}}_{\alpha}$

ρ \ Q	.2	.3	.4	.5	.6	.7	.8	.9
.5	0.00	0.00	0.83	7.14	2.85	0.99	0.25	0.02
.6	0.82	2.96	6.39	11.68	8.19	5.95	4.11	2.23
.7	7.06	10.49	14.20	18.70	16.02	13.28	9.98	5.71
.8	15.06	20.89	25.81	30.18	28.69	25.49	20.15	12.08
.9	25.60	36.04	44.65	51.26	53.03	50.73	43.08	27.97

The percent gains of both estimates increase as ρ increases and Q approaches .5.

5.3.2. Properties of ${}_2\overline{\overline{Y}}_{\alpha_2}$ when $Q \leq \frac{1}{2}$

Let us start from $\alpha = 3$ to derive the optimum estimator ${}_2\overline{\overline{Y}}_{\alpha_2 w}$ and its variance when $Q \leq \frac{1}{2}$.

$(\alpha = 3)$

An estimator of $\overline{\overline{Y}}_3$ which belongs to the modified composite class of estimators is

$$ {}_2\bar{Y}_3 = a_3({}_2\bar{Y}_2 + \bar{Y}_{31} - \bar{Y}_{23}) + (1-a_3)\bar{Y}_{32} \quad , \qquad (5.23)$$

or,

$$ {}_2\bar{Y}_3 = \bar{Y}_{32} + a_3(\bar{Y}_{31}-\bar{Y}_{32}) + a_3a_2(\bar{Y}_{21}-\bar{Y}_{22}) + a_3a_2Q(\bar{Y}_{11}-\bar{Y}_{12})$$

$$ + a_3(\bar{Y}_{22} - \bar{Y}_{23}) \quad .$$

Figure 5.3 is drawn according to the formation of ${}_2\bar{Y}_\alpha$ when $\frac{1}{3} \le Q \le \frac{1}{2}$.

<u>Variance of ${}_2\bar{Y}_3$:</u>

$$V({}_2\bar{Y}_3) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + a_3^2 \frac{N}{N-1}\frac{\sigma_3^2}{nPQ} + a_2^2a_2^2 \frac{N}{N-1}\frac{\sigma_2^2}{nPQ} + a_2^2a_2^2Q^2 \frac{N}{N-1}\frac{\sigma_1^2}{nPQ}$$

$$- 2a_3 \frac{N}{N-1}\frac{\sigma_3^2}{Qn} - 2\, a_3^2a_2 \frac{N}{N-1}\frac{\sigma_{23}}{nPQ}\frac{Q^2}{P} - 2a_2^2a_2Q^2 \frac{N}{N-1}\frac{\sigma_{13}}{nPQ}\frac{P-Q}{P}$$

$$- 2\, a_3^2a_2^2Q^2 \frac{N}{N-1}\frac{\sigma_{12}}{nPQ} + a_3^2 \frac{N}{N-1}\frac{P-Q}{nPQ}\sigma_2^2 - 2\, a_3^2a_2 \frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\frac{P-Q}{P}$$

$$+ 2\, a_3^2a_2Q^2 \frac{N}{N-1}\frac{\sigma_{12}}{nPQ}\frac{P-Q}{P} \quad .$$

The derivation is given in Appendix 9.8.

<u>Optimum estimator and its variance ($\alpha = 3$)</u>: The optimum values of a_i can be found as follows;

$$a_{2w} = \frac{Q^2\sigma_{23} + (P-Q)\sigma_2^2 + Q^2(P-Q)\sigma_{13} - Q^2(P-Q)\sigma_{12}}{P(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12})}$$

$$\frac{(N-1)}{2}\frac{nPQ}{N}\frac{\partial}{\partial a_3} V({}_2\bar{Y}_3) = a_3\left[\sigma_3^2 + a_2^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}) - 2\frac{a_2}{P}(Q^2\sigma_{23}\right.$$

$$\left. + (P-Q)\sigma_2^2 + Q^2(P-Q)\sigma_{13} - Q^2(P-Q)\sigma_{12}) + (P-Q)\sigma_2^2\right] - P\sigma_3^2 = 0 \quad .$$

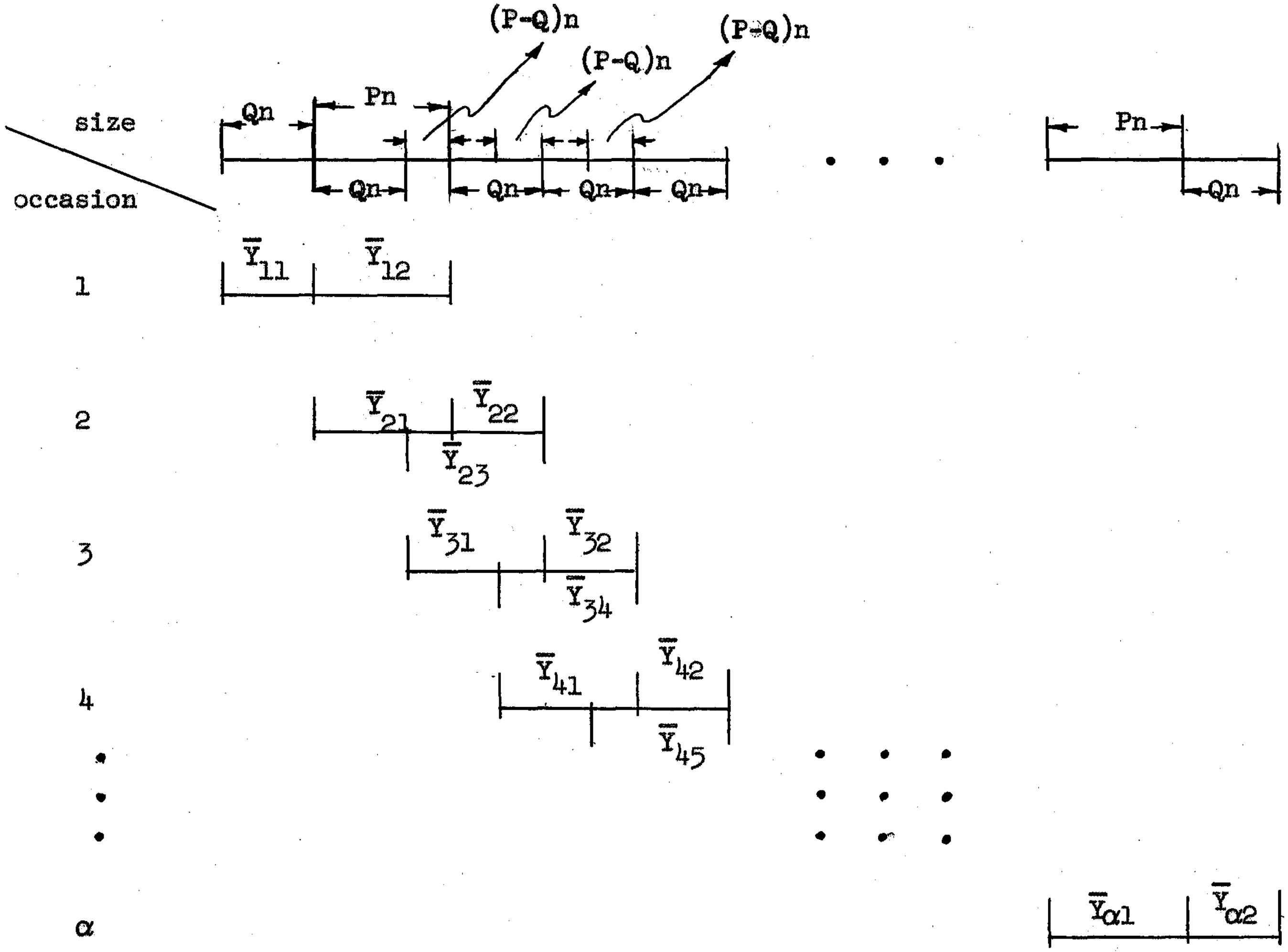


Figure 5.3. Formation of ${}_2\bar{Y}_{\alpha.}$ when $Q \leq \frac{1}{2}$

From a_{2w} ,

$$a_{3w} = \frac{P\sigma_3^2}{\sigma_3^2 - a_{2w}^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}) + (P-Q)\sigma_2^2} \qquad . \qquad (5.24)$$

This provides the optimum estimator ${}_2\overline{Y}_{3w}$ of $\overline{\overline{Y}}_3$ when $\frac{1}{3} \leq Q \leq \frac{1}{2}$, thus $V({}_2\overline{Y}_3)$ can be rewritten as

$$V({}_2\overline{Y}_3) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + a_3^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\Big[\sigma_3^2 + a_2^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12})$$

$$+ (P-Q)\sigma_2^2 - 2\,\frac{a_2}{P}\left(Q^2\sigma_{23} + Q^2(P-Q)\sigma_{13} + (P-Q)\sigma_2^2 - Q^2\,\frac{(P-Q)}{P}\,\sigma_{12}\right)\Big]$$

$$- 2\,Pa_3\,\frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ} \qquad .$$

From a_{2w}, a_{3w} , we have

$$V({}_2\overline{Y}_{3w}) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + a_{3w}^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\Big[\sigma_3^2 - a_{2w}^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12})$$

$$+ (\Sigma-Q)\sigma_2^2\Big] - 2Pa_{3w}\,\frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}$$

$$= \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}\,P\,a_{3w}$$

$$= \frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}\,(P - Pa_{3w}) - \frac{\sigma_3^2}{N-1} \qquad .$$

When $Q = \frac{1}{2}$, this result is identical with the result of $V({}_2\overline{Y}_{3w})$ in (5.16) when $\frac{1}{2} \leq Q < 1$. Hence

$$V({}_2\overline{Y}_{\alpha_2 w}) = \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,(P - P\,a_{\alpha w}) - \frac{\sigma_\alpha^2}{N-1} \qquad . \qquad (5.25)$$

The percent gain of ${}_2\overline{Y}_{\alpha_2}$ over $\overline{Y}_\alpha$ when $Q \leq \frac{1}{2}$ is given in Table 5.5. It can be seen from the table that the percent gain is decreasing rapidly when $Q \leq \frac{1}{2}$.

Table 5.6. Comparison of ${}_2\overline{Y}_{\alpha_i}$ with ${}_5\overline{Y}_{\alpha_i}$

Q = .5

ρ	M_1/S_1	M_3/S_2	M_3/S_3	M_4/S_4
.5	1.0714	1.0051	1.0004	1.0000
.6	1.0513	1.0057	1.0006	1.0001
.7	1.0331	1.0056	1.0009	1.0002
.8	1.0173	1.0045	1.0011	1.0003
.9	1.0052	1.0022	1.0009	1.0003
Q = .6				
.5	1.0666	1.0401	1.0389	1.0388
.6	1.0439	1.0242	1.0228	1.0227
.7	1.0177	1.0016	.9994	.9992
.8	.9848	.9631	.9575	.9563
.9	.9403	.8859	.8622	.8535
Q = .7				
.5	1.0567	1.0477	1.0475	1.0475
.6	1.0342	1.0268	1.0265	1.0265
.7	1.0035	.9950	.9942	.9942
.8	.9581	.9395	.9365	.9361
.9	.8863	.8283	.8119	.8077
Q = .8				
.5	1.0421	.9600	.9600	.9600
.6	1.0230	.9789	.9790	.9790
.7	.9934	.9900	.9898	.9898
.8	.9432	.9329	.9320	.9320
.9	.8517	.8113	.8043	.8032
Q = .9				
.5	1.0231	.9771	.9771	.9771
.6	1.0114	.9888	.9888	.9888
.7	.9906	.9900	.9900	.9900
.8	.9500	.9475	.9474	.9474
.9	.8591	.8439	.8428	.8427

5.4. Comparison of a Symmetric Estimator and a Modified Composite Estimator

From a comparison of Table 5.3 and Table 5.5, it can be seen that the symmetric estimator over i = 1.2 is better than the modified composite estimator when ρ and Q are relatively small.

Table 5.6 gives another comparison between these two estimators by using a ratio of the two variances. From Table 5.6, the same conclusion as the above is obtained. That is, the symmetric estimator is better than the modified composite estimator when ρ and Q are small. However, another interesting fact is that the relative efficiency of the modified composite estimator increases as the number of previous occasions increases. Even when Q is .5, both estimators have the same precision if the number of previous occasions used in estimating $\bar{\bar{Y}}_\alpha$ is more than three.

6. ESTIMATION THEORY FOR A COMPLETE ESTIMATOR

All studies in the previous chapters are based on the proposed sampling scheme described in 3.1. In fact, the aim of a partial retention policy in successive sampling is to improve an estimator by using previous information in estimating the current population mean. Based on this general principle, a new alternative sampling scheme will be considered in this chapter.

It has already been observed that the general linear estimator is most efficient on two occasions. This estimator was formed in the following way. The sample of $n + Qn$ units is partitioned into disjoint sets and the estimator is a linear combination of sample means based on those disjoint sets. To extend this idea for more than two occasions, a new sampling scheme has been devised. In 6.1, this new sampling scheme and a retention policy based on this scheme are introduced. In 6.2 the complete estimator based on this new scheme and its properties will be discussed. In 6.3 comparisons of this new estimator with the two previous estimators will be examined.

6.1. A New Sampling Scheme

The new sampling scheme retains Pn first-stage units from the first occasion in the sample for every successive occasion; the remaining Qn units in the sample are replaced on every occasion.

The sampling scheme and the retention scheme are defined as follows.

Assume the population U consists of N first stage units on α occasions: $\{u_i\}$, $i = 1, 2, \ldots, N$. In each u_i, there are N_i second-stage units.

1. As with the proposed sampling scheme in 3.1, draw from U a preliminary first-stage sample of size $n + (\alpha-1)Qn$ with equal probabilities without replacement of units at each draw. Record the order of each unit.

2. The first stage units which occurred from order 1 to n constitute the sample for the first occasion. Then draw n_i second-stage units as in the previous sampling scheme described in 3.1.

3. Retain the first Pn units for the second occasion and reject the next Qn units, supplementing the retained units by the next set of Qn units which occurred from order n+1 to n+Qn. Thus the sample size of n first stage units is maintained on the second occasion, having the first Pn units matched with those of the first occasion. The second stage units originally selected from the matched Pn first stage units are also retained for the second occasion. On the third occasion retain again the same first Pn units and reject the Qn units which were used as replacements on the second occasion. Supplement the retained units by Qn units which occurred from n+Qn+1 to n+2Qn in the sample selection. Thus the first Pn units are matched with those of the first and second occasions. Repeat this procedure for the succeeding occasions. On the αth occasion, there will be retained the first Pn units, which are matched with those of all previous occasions, and Qn unmatched units which occurred from order $n+(\alpha-2)Qn+1$ to $n+(\alpha-1)Qn$. Figure 6.1 illustrates this scheme.

6.2. A Complete Estimator and Its Properties

According to the new sampling scheme, the complete estimator of $\bar{\bar{Y}}_\alpha$, denoted by ${}_7\bar{\bar{Y}}_\alpha$, is defined as a linear combination of sample means which are based on disjoint sets on each ith occasion.

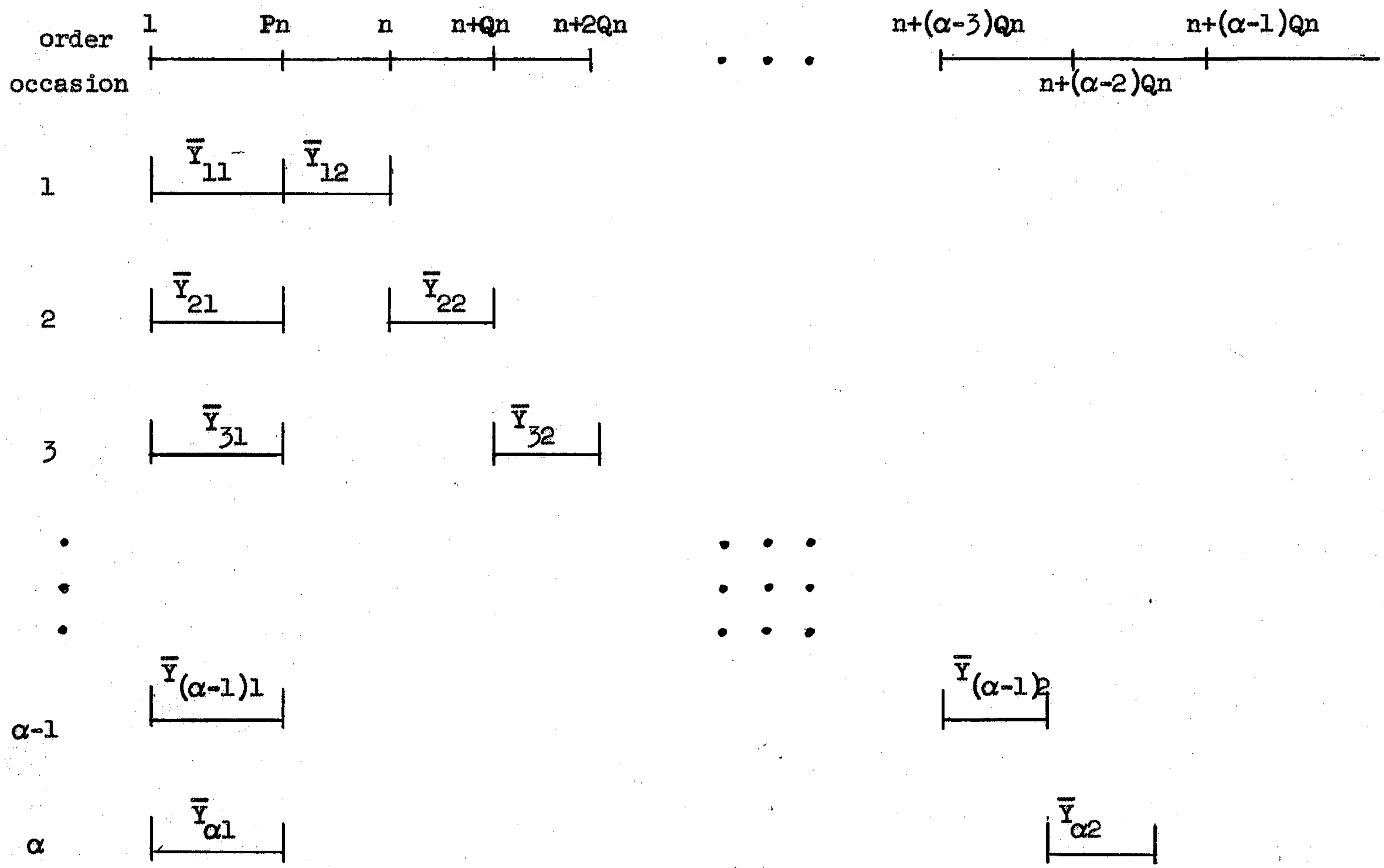


Figure 6.1. New sampling scheme

$$_7\bar{Y}_{\alpha.} = \sum_{i=o}^{\alpha-1} a_i \bar{Y}_{(\alpha-i)1} + \Sigma b_i \bar{Y}_{(\alpha-i)2} ,$$

where $a_o + b_o = 1$ $a_i + b_i = 0$ for all $i = 1, 2, \ldots, \alpha-1$,

or

$$_7\bar{Y}_{\alpha.} = \bar{Y}_{\alpha 2} + \sum_{i=o}^{\alpha-1} a_i (\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}) . \tag{6.1}$$

Hence, its variance is

$$V(_7\bar{Y}_{\alpha.}) = V(\bar{Y}_{\alpha.2}) + \sum_{i=o}^{\alpha-1} a_i V(\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2})$$

$$+ 2a_o \operatorname{cov} (\bar{Y}_{\alpha 2}, \bar{Y}_{\alpha 1} - \bar{Y}_{\alpha 2}) + 2 \Sigma a_i a_j \operatorname{cov} (\bar{Y}_{(\alpha-i)1} - \bar{Y}_{(\alpha-i)2}, \bar{Y}_{(\alpha-j)1}$$

$$- \bar{Y}_{(\alpha-j)2}) = \frac{\sigma_\alpha^2}{Pn} \frac{N-Pn}{N-1} + \sum_{i=o}^{\alpha-1} a_i^2 \frac{N}{N-1} \frac{\sigma_{\alpha-i}^2}{nPQ} - 2a_o \frac{N}{N-1} \frac{\sigma_\alpha^2}{Qn}$$

$$+ 2 \sum_{i(<j)=o}^{\alpha-2} a_i a_j \frac{N}{N-1} \frac{\sigma_{(\alpha-1)(\alpha-j)}}{Pn} \tag{6.2}$$

From Lemma 1, this estimator is an unbiased estimator of $\bar{\bar{Y}}_\alpha$.

<u>Optimum estimator and its variance</u>: In order to find the optimum estimator of $\bar{\bar{Y}}_\alpha$ using all α-1 previous occasions, let us start from the simplest case. When α=2 the optimum estimator $_7\bar{Y}_2$ is the same as $_1\bar{Y}_2$ in 4.2.1.

(α=3)

The estimator is

$$_7\bar{Y}_3 = a_o\bar{Y}_{31} + b_o\bar{Y}_{32} + a_1\bar{Y}_{21} + b_1\bar{Y}_{22} + a_2\bar{Y}_{11} + b_2\bar{Y}_{12}$$

$$= \bar{Y}_{32} + a_o(\bar{Y}_{31} - \bar{Y}_{32}) + a_1(\bar{Y}_{21} - \bar{Y}_{22}) + a_2 (\bar{Y}_{11} - \bar{Y}_{12}) .$$

Hence,

$$V({}_7\bar{Y}_3) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + a_o^2 \frac{N}{N-1}\frac{\sigma_3^2}{nPQ} + a_1^2 \frac{N}{N-1}\frac{\sigma_2^2}{nPQ} + a_2^2 \frac{N}{N-1}\frac{\sigma_1^2}{nPQ}$$

$$- 2\, a_o \frac{N}{N-1}\frac{\sigma_3^2}{Qn} + 2a_o a_1 \frac{N}{N-1}\frac{\sigma_{23}}{Pn} + 2a_o a_2 \frac{N}{N-1}\frac{\sigma_{13}}{Pn} + 2a_1 a_2 \frac{N}{N-1}\frac{\sigma_{12}}{Pn} .$$

(6.3)

The equation for the optimum value of a_i is

$$A_3\, \underline{a} = \underline{c} \quad ,$$

where

$$A_3 = \begin{bmatrix} \sigma_1^2 & Q\sigma_{12} & Q\sigma_{13} \\ Q\sigma_{12} & \sigma_2^2 & Q\sigma_{23} \\ Q\sigma_{13} & Q\sigma_{23} & \sigma_3^2 \end{bmatrix} \quad \underline{a} = \begin{bmatrix} a_2 \\ a_1 \\ a_o \end{bmatrix} \quad \underline{c} = \begin{bmatrix} 0 \\ 0 \\ P\sigma_3^2 \end{bmatrix} .$$

Let us denote the cofactor of an element of the ith row and jth column in A_3 by A_{ij}, and also denote the minor determinant of an element of the ith row and jth column in the corresponding correlation matrix of A_3 by Δ_{ij}. Expressing explicitly,

$$A_{31} = \sigma_2^2\sigma_1\sigma_3 \begin{vmatrix} Q\rho_{12} & Q\rho_{13} \\ 1 & Q\rho_{23} \end{vmatrix} \qquad \Delta_{31} = \begin{vmatrix} Q\rho_{12} & Q\rho_{13} \\ 1 & Q\rho_{23} \end{vmatrix}$$

$$A_{32} = -\sigma_1^2\sigma_2\sigma_3 \begin{vmatrix} 1 & Q\rho_{13} \\ Q\rho_{12} & Q\rho_{23} \end{vmatrix} \qquad \Delta_{32} = \begin{vmatrix} 1 & Q\rho_{13} \\ Q\rho_{12} & Q\rho_{23} \end{vmatrix}$$

$$A_{33} = \sigma_1^2\sigma_2^2 \begin{vmatrix} 1 & Q\rho_{12} \\ Q\rho_{12} & 1 \end{vmatrix} \qquad \Delta_{33} = \begin{vmatrix} 1 & Q\rho_{12} \\ Q\rho_{12} & 1 \end{vmatrix}$$

$$|A_3| = \sigma_1^2\sigma_2^2\sigma_3^2 \begin{vmatrix} 1 & Q\rho_{12} & Q\rho_{13} \\ Q\rho_{12} & 1 & Q\rho_{23} \\ Q\rho_{13} & Q\rho_{23} & 1 \end{vmatrix} \qquad \Delta_3 = \begin{vmatrix} 1 & Q\rho_{12} & Q\rho_{13} \\ Q\rho_{12} & 1 & Q\rho_{23} \\ Q\rho_{13} & Q\rho_{23} & 1 \end{vmatrix} .$$

Then, the optimum value of a_i is

$$\begin{aligned} a_{2w} &= \frac{1}{\sigma_1\sigma_3}\,\frac{\Delta_{31}}{\Delta_3}\,P\sigma_3^2 \\ a_{1w} &= -\frac{1}{\sigma_2\sigma_3}\,\frac{\Delta_{32}}{\Delta_3}\,P\sigma_3^2 \\ a_{ow} &= \frac{1}{\sigma_3^2}\,\frac{\Delta_{33}}{\Delta_3}\,P\sigma_3^2 \quad . \end{aligned} \tag{6.4}$$

Hence, the variance of the optimum estimator ${}_7\bar{\bar{Y}}_{3w}$ is obtained as

$$V({}_7\bar{\bar{Y}}_{3w}) = \frac{\sigma_3^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}\,\frac{P^2}{\Delta_3^2}\left(\Delta_{33}^2 - \Delta_{32}^2 - \Delta_{31}^2 + 2Q\rho_{12}\Delta_{13}\Delta_{23}\right) .$$

From this, the variance of the optimum estimator ${}_7\bar{\bar{Y}}_{\alpha_2 w}$, using two previous occasions in estimating $\bar{\bar{Y}}_\alpha$, can be reduced to

$$\begin{aligned} V({}_7\bar{\bar{Y}}_{\alpha_2 w}) &= \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{P^2}{\Delta_3^2}\left(\Delta_{33}^2 - \Delta_{32}^2 - \Delta_{31}^2\right. \\ &\quad \left. + 2Q\rho_{(\alpha-2)(\alpha-1)}\Delta_{31}\Delta_{32}\right) \\ &= \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,P^2\psi_2 \end{aligned} \tag{6.5}$$

where

$$\psi_2 = \frac{1}{\Delta_3^2}\left(\Delta_{33}^2 - \Delta_{32}^2 - \Delta_{31}^2 + 2Q\rho_{(\alpha-2)(\alpha-1)}\Delta_{31}\Delta_{32}\right)$$

and the correlations in ψ_2 are defined appropriately.

(α=4)

The estimator is

$$_7\bar{Y}_4 = \bar{Y}_{42} + a_o(\bar{Y}_{41} - \bar{Y}_{42}) + a_1(\bar{Y}_{31} - \bar{Y}_{32}) + a_2(\bar{Y}_{21} - \bar{Y}_{22}) + a_3(\bar{Y}_{11} - \bar{Y}_{12}),$$

and its variance is

$$V(_7\bar{Y}_4) = \frac{\sigma_4^2}{Qn}\left(\frac{N-Qn}{N-1}\right) + \sum_{i=o}^{3} a_i^2 \frac{N}{N-1}\frac{\sigma_{4-i}^2}{nPQ} - 2a_o \frac{N}{N-1}\frac{\sigma_4^2}{Qn}$$

$$+ 2\sum_{i<j}^{2} a_i a_j \frac{N}{N-1}\frac{\sigma_{(4-i)(4-j)}}{Pn} \quad . \tag{6.6}$$

The equation for the optimum a_i is

$$A_4 \underline{a} = \underline{c} \quad ,$$

where

$$A_4 = \begin{bmatrix} \sigma_1^2 & Q\sigma_{12} & Q\sigma_{13} & Q\sigma_{14} \\ Q\sigma_{12} & \sigma_2^2 & Q\sigma_{23} & Q\sigma_{24} \\ Q\sigma_{31} & Q\sigma_{32} & Q\sigma_3^2 & Q\sigma_{34} \\ Q\sigma_{41} & Q\sigma_{42} & Q\sigma_{43} & \sigma_4^2 \end{bmatrix} \quad \underline{a} = \begin{bmatrix} a_3 \\ a_2 \\ a_1 \\ a_o \end{bmatrix} \quad \underline{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ P\sigma_4^2 \end{bmatrix} \quad .$$

Hence, the optimum values of the a_i are obtained as follows.

$$a_{3w} = -\frac{\sigma_4}{\sigma_1} P \frac{\Delta_{41}}{\Delta_4}$$

$$a_{2w} = \frac{\sigma_4}{\sigma_1} P \frac{\Delta_{42}}{\Delta_4} \tag{6.7}$$

$$a_{1w} = -\frac{\sigma_4}{\sigma_1} P \frac{\Delta_{43}}{\Delta_4}$$

$$a_{ow} = P \frac{\Delta_{44}}{\Delta_4} \quad ,$$

where Δ_{ij} s are defined as in $\alpha = 3$.

The variance of the optimum estimator of ${}_7\overline{Y}_{4w}$ can be obtained as follows.

$$V({}_7\overline{Y}_{4w}) = \frac{\sigma_4^2}{Qn} - \frac{N}{N-1}\frac{\sigma_4^2}{nPQ}\frac{P^2}{\Delta_4^2}(\Delta_{44}^2 + \Delta_{43}^2 - \Delta_{42}^2 - \Delta_{41}^2$$

$$- 2Q\rho_{34}\Delta_{44}\Delta_{43} + 2Q\rho_{12}\Delta_{41}\Delta_{42})$$

$$= \frac{\sigma_4^2}{Qn} - \frac{N}{N-1}\frac{\sigma_4^2}{nPQ}P^2\psi_3 \quad .$$

Hence, the variance of ${}_7\overline{Y}_{\alpha_3 w}$ is, from $V({}_7\overline{Y}_{4w})$,

$$V({}_7\overline{Y}_{\alpha_3 w}) = \frac{\sigma_\alpha^2}{Qn} - \frac{N}{N-1}\frac{\sigma_\alpha^2}{nPQ}P^2\psi_3 \quad , \qquad (6.8)$$

where the correlations in ψ_3 are defined appropriately.

$(\alpha=5)$

The estimator is

$${}_7\overline{Y}_5 = \overline{Y}_{52} + \sum_{i=o}^{4} a_i\,(\overline{Y}_{(5-i)1} - \overline{Y}_{(5-i)2})$$

and its variance can be obtained from (6.2).

By similar methods, the optimum values of a_i are obtained as follows.

$$a_{4w} = \frac{\sigma_5}{\sigma_1}P\frac{\Delta_{51}}{\Delta_5}$$

$$a_{3w} = -\frac{\sigma_5}{\sigma_2}P\frac{\Delta_{52}}{\Delta_5}$$

$$a_{2w} = \frac{\sigma_5}{\sigma_3}P\frac{\Delta_{53}}{\Delta_5} \qquad (6.9)$$

$$a_{1w} = -\frac{\sigma_5}{\sigma_4}P\frac{\Delta_{54}}{\Delta_5}$$

$$a_{ow} = P\,\frac{\Delta_{55}}{\Delta_5}$$

and the variance of the optimum estimator $_7\bar{Y}_{5w}$ is

$$V(_7\bar{Y}_{5w}) = \frac{\sigma_5^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_5^2}{nPQ}\,\frac{P^2}{\Delta_5^2}\,(\Delta_{55}^2 + \Delta_{54}^2 - \Delta_{53}^2$$

$$- \Delta_{52}^2 - \Delta_{51}^2 - 2Q\rho_{45}\Delta_{54}\Delta_{55} + 2Q\rho_{23}\Delta_{52}\Delta_{53}$$

$$- 2Q\rho_{13}\Delta_{51}\Delta_{53} + 2Q\rho_{12}\Delta_{51}\Delta_{52}) \;.$$

Hence, the variance of $_7\bar{Y}_{\alpha_4 w}$ is

$$V(_7\bar{Y}_{\alpha_4 w}) = \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\quad \psi_4 \quad . \tag{6.10}$$

$(\alpha=\alpha)$

The estimator and its variance are given in (6.1) and (6.2). The optimum value of a_i is as follows.

$$a_{\alpha-1w} = (-1)^{\alpha-1}\,\frac{1}{\sigma_1\sigma_\alpha}\,\frac{\Delta_{\alpha 1}}{\Delta_\alpha}\,P\,\sigma_\alpha^2$$

$$a_{\alpha-2w} = (-1)^{\alpha-2}\,\frac{1}{\sigma_\alpha\sigma_\alpha}\,\frac{\Delta_{\alpha 2}}{\Delta_\alpha}\,P\,\sigma_\alpha^2$$

$$\vdots$$

$$a_{iw} = (-1)^{i}\,\frac{1}{\sigma_{\alpha-i}\sigma_\alpha}\,\frac{\Delta_{\alpha(\alpha-i)}}{\Delta_\alpha}\,P\,\sigma_\alpha^2$$

$$\vdots$$

$$a_{2w} = \frac{1}{\sigma_{\alpha-2}\sigma_\alpha}\,\frac{\Delta_{\alpha(\alpha-2)}}{\Delta_\alpha}\,P\,\sigma_\alpha^2$$

$$a_{1w} = -\frac{1}{\sigma_{\alpha-1}\sigma_\alpha}\,\frac{\Delta_{\alpha(\alpha-1)}}{\Delta_\alpha}\,P\,\sigma_\alpha^2$$

$$a_{ow} = \frac{\Delta_{\alpha\alpha}}{\Delta_\alpha}\,P$$

and the variance of ${}_7\bar{Y}_{\alpha.w}$ is

$$V({}_7\bar{Y}_{\alpha.w}) = \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ}\,\frac{P^2}{\Delta_\alpha^2}\Big(\Delta_{\alpha\alpha}^2 - \sum_{i=1}^{\alpha-1}\Delta_{\alpha(\alpha-i)}^2$$

$$+ (-1)^{i+j-1}\, 2 \sum_{i=1}^{\alpha-2}\sum_{j=i+1}^{\alpha-1} Q\rho_{(\alpha-i)(\alpha-j)}\Delta_{\alpha(\alpha-i)}\Delta_{\alpha(\alpha-j)}\Big) .$$

<u>Relationship of percent loss of precision to number of previous occasions included in estimator</u>: In the following numerical analysis, the assumptions are $\sigma_{\alpha-i}^2 = \sigma^2$ and $\rho_{(\alpha-i)(\alpha-i+1)} = \rho$, $\rho_{(\alpha-i)(\alpha-i+2)} = \rho^2$, ..., for all i . Table 6.1 provides the percent gain of an estimator ${}_7\bar{Y}_{\alpha_i}$ when i previous occasions are used in estimating $\bar{\bar{Y}}_\alpha$.

In Table 6.1, percent loss of precision decreases rapidly from I_1 to I_2 and it is less than 5% when the number of previous occasions is two. From this analysis, it may be inferred that the preferred number of previous occasions is two, over all values of Q larger than .2 .

In Table 6.1, the optimum weights of the estimator using two previous occasions in estimating $\bar{\bar{Y}}_5$ are given. The values of a_{ow} for $\bar{Y}_{51}$ are relatively high, and the values of a_{iw} for previous occasions rapidly decrease. The optimum value a_{ow} increases as the matching fraction P and correlation coefficient increase.

Next, percent gain of ${}_7\bar{Y}_{\alpha_1}$ and ${}_7\bar{Y}_{\alpha_2}$ over $\bar{Y}_\alpha$ is presented in Table 6.2.

The percent gain is directly related to the correlation coefficient and it is observed that for both cases the maximum value of the gain is attained when Q is near .5.

Table 6.1. Percent loss of precision of ${}_7\overline{Y}_{\alpha_i}$ and optimum values of a_i in ${}_7\overline{Y}_{\alpha_2}$

Q = .2

ρ \ I_i	I_1	I_2	I_3	ρ \ A_{iw}	A_{0w}	$-A_{1w}$	$-A_{2w}$
.5	2.73	.13	.01	.5	.809	.077	.032
.6	4.73	.44	.08	.6	.814	.092	.047
.7	8.18	1.31	.32	.7	.821	.105	.065
.8	14.56	3.58	1.10	.8	.830	.118	.087
.9	27.75	9.65	3.58	.9	.841	.131	.112

Q = .3

.5	3.27	.10	.01	.5	.718	.101	.038
.6	5.55	.37	.05	.6	.727	.120	.056
.7	9.48	1.14	.24	.7	.740	.138	.079
.8	16.90	3.35	.91	.8	.758	.156	.108
.9	33.45	9.09	3.43	.9	.781	.172	.143

Q = .4

.5	3.41	.07	.00	.5	.627	.117	.039
.6	5.67	.25	.03	.6	.642	.139	.058
.7	9.53	.84	.14	.7	.661	.161	.084
.8	16.98	2.66	.63	.8	.688	.182	.117
.9	34.82	9.08	2.81	.9	.725	.202	.162

Q = .5

.5	3.24	.03	.00	.5	.535	.125	.035
.6	5.26	.15	.01	.6	.554	.150	.054
.7	8.68	.52	.07	.7	.581	.175	.081
.8	15.41	1.83	.36	.8	.617	.200	.117
.9	32.65	7.24	2.01	.9	.670	.225	.170

(Table 6.1 continued)

(Table 6.1 continued)

Q = .6

ρ \ I_i	I_1	I_2	I_3
.5	2.63	.01	.00
.6	3.93	.07	.00
.7	5.92	.27	.02
.8	9.29	1.06	.18
.9	18.51	5.05	1.22

ρ \ A_{iw}	A_{0w}	$-A_{1w}$	$-A_{2w}$
.5	.441	.123	.029
.6	.464	.150	.046
.7	.495	.178	.070
.8	.541	.207	.108
.9	.610	.239	.167

Q = .7

.5	2.26	.00	.00
.6	3.46	.02	.00
.7	5.38	.10	.00
.8	9.24	.48	.05
.9	20.69	2.89	.57

.5	.343	.112	.020
.6	.367	.140	.033
.7	.402	.170	.054
.8	.454	.204	.088
.9	.540	.244	.152

Q = .8

.5	1.57	.00	.00
.6	2.31	.00	.00
.7	3.42	.02	.00
.8	5.59	.13	.00
.9	12.59	1.13	.16

.5	.238	.090	.011
.6	.261	.116	.019
.7	.295	.146	.033
.8	.349	.184	.060
.9	.447	.235	.120

Q = .9

.5	0.81	.00	.00
.6	1.13	.00	.00
.7	1.57	.00	.00
.8	2.30	.01	.00
.9	4.74	.15	.01

.5	.125	.054	.003
.6	.141	.072	.006
.7	.166	.097	.012
.8	.210	.133	.025
.9	.304	.194	.064

Table 6.2. Percent gain of ${}_7\overline{Y}_{\alpha_1}$, ${}_7\overline{Y}_{\alpha_2}$ over $\overline{Y}_\alpha$

${}_7\overline{Y}_{\alpha_1}$ over $\overline{Y}_\alpha$

ρ \ Q	.2	.3	.4	.5	.6	.7	.8	.9
.5	4.21	5.67	6.66	7.14	7.05	6.36	4.99	2.90
.6	6.20	8.47	10.09	10.97	11.03	10.10	8.08	4.79
.7	8.69	12.06	14.53	16.22	16.65	15.66	12.89	7.88
.8	11.74	16.63	20.64	23.52	25.93	24.34	20.98	13.58
.9	15.46	22.47	28.75	34.03	37.82	39.28	36.81	26.89

${}_7\overline{Y}_{\alpha_2}$ over $\overline{Y}_\alpha$

ρ \ Q	.2	.3	.4	.5	.6	.7	.8	.9
.5	4.93	6.45	7.36	7.69	7.42	6.56	5.07	2.91
.6	7.79	10.24	11.75	12.32	11.96	10.64	8.30	4.82
.7	11.88	15.77	18.27	19.36	18.99	17.08	13.51	8.00
.8	17.77	24.07	28.45	30.76	30.80	28.31	22.93	14.01
.9	26.49	37.20	45.73	51.59	54.09	52.31	44.84	29.39

6.3. Comparison of a Complete Estimator with Two Previous Estimators

The comparison of the estimators will be achieved in two parts. The first part is the comparison of the complete estimator with the two previous estimators for i from 1 to 4 and $Q \geq \frac{1}{2}$. The second part is the comparison of the complete estimator with the modified composite estimator over all Q when $i = 2$. This is because it was determined in 6.2 that the preferred number of previous occasions for the complete estimator is two, for all values of Q.

The first part, the comparison of the complete estimator with two previous estimators, can be made from the tables of the percent gains of the estimators over $\overline{Y}_\alpha$ in Table 5.3, 5.5 and 6.2. From these comparisons, it can be seen that the complete estimator is best among the three estimators when ρ and Q are larger than .5. Based on these tables, Figure 6.2 shows the precision of ${}_2\bar{Y}_{\alpha_2}$, ${}_5\bar{Y}_{\alpha_2}$ and ${}_7\bar{Y}_{\alpha_2}$ relative to $\bar{Y}_\alpha$.

Tables 6.3 and 6.4 provide another comparison of the complete estimator with the other two estimators by means of ratios of variances of the estimators. The ratios are defined as follows, ignoring $\frac{\sigma_\alpha^2}{N-1}$ in the variance of each estimator:

$$\frac{V({}_5\bar{Y}_{\alpha_i})}{V({}_7\bar{Y}_{\alpha_i})} = \frac{S_i}{C_i}$$

$$\frac{V({}_2\bar{Y}_{\alpha_i})}{V({}_7\bar{Y}_{\alpha_i})} = \frac{M_i}{C_i} ,$$

where $C_i = P - P^2 \psi_i$ in (6.5),(6.8), (6.10) .

From Table 6.3, for all i, the complete estimator is better than the symmetric estimator for values of ρ and Q larger than .5. It is interesting to note that when Q = .5, the two optimum estimators have exactly the same precision regardless of the number of previous occasions. This means that when Q = .5 the general linear estimator based on the proposed sampling scheme in 3.1 has exactly the same precision as the complete estimator based on the new sampling scheme, regardless of the number of previous occasions used in estimating the current population mean $\bar{\bar{Y}}_\alpha$. Also, the ratio increases as ρ and i , the number of previous

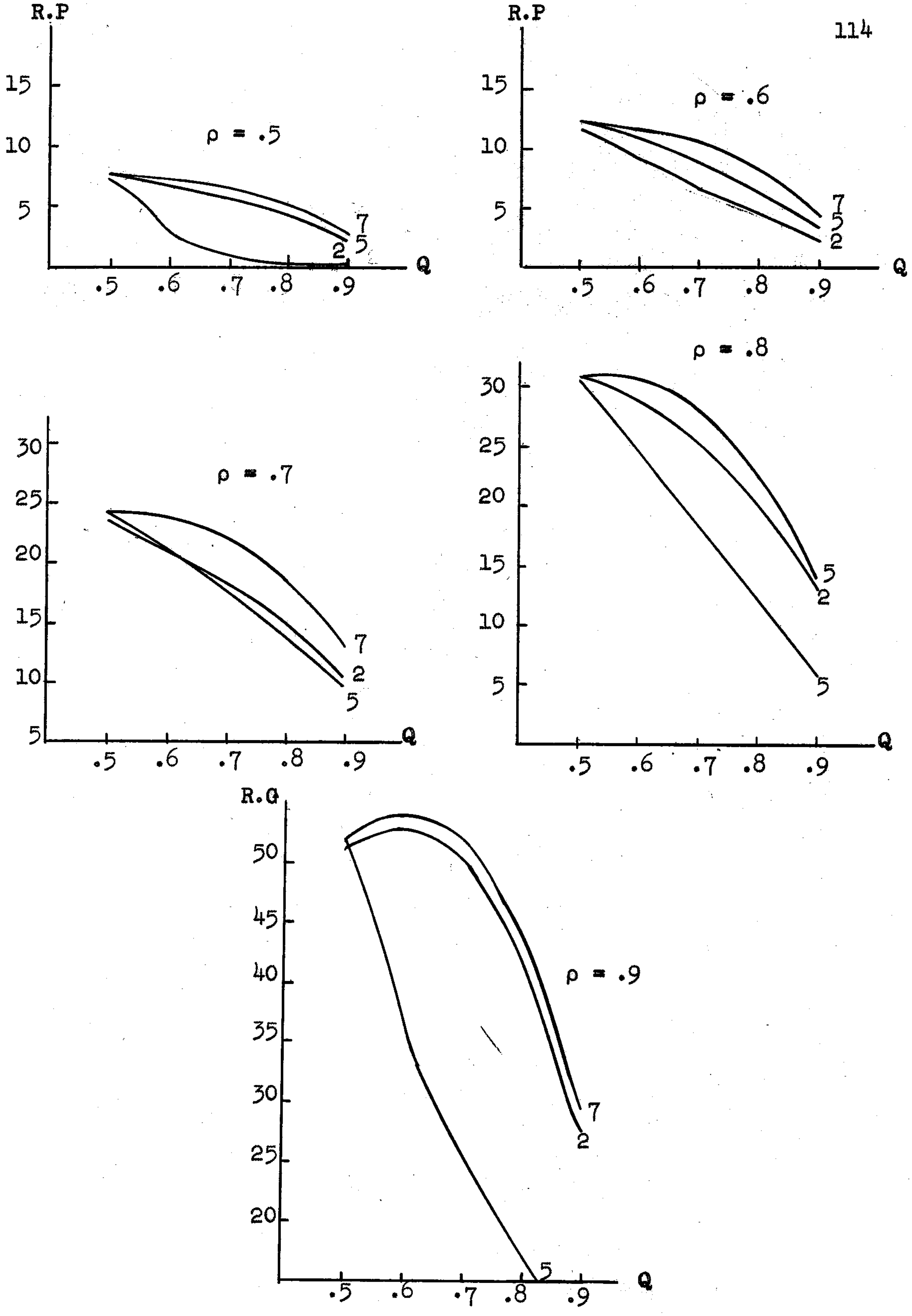


Figure 6.2. Precision of the estimators ${}_2\overline{Y}_{\alpha_2}$, ${}_5\overline{Y}_{\alpha_2}$ and ${}_7\overline{Y}_{\alpha_2}$ relative to $\overline{Y}_\alpha$

Table 6.3. Comparison of ${}_5\bar{Y}_{\alpha_i}$ with ${}_7\bar{Y}_{\alpha_i}$

Q = .5

ρ	S_1/C_1	S_2/C_2	S_3/C_3	S_4/C_4
.5	1.00	1.00	1.00	1.00
.6	1.00	1.00	1.00	1.00
.7	1.00	1.00	1.00	1.00
.8	1.00	1.00	1.00	1.00
.9	1.00	1.00	1.00	1.00
		Q = .6		
.5	1.00067	1.0042	1.0042	1.0042
.6	1.00854	1.0103	1.0105	1.0105
.7	1.01774	1.00239	1.0251	1.0253
.8	1.03550	1.0554	1.0610	1.0623
.9	1.06811	1.1365	1.1684	1.1806
		Q = .7		
.5	1.00650	1.0070	1.0071	1.0071
.6	1.01503	1.0170	1.0171	1.0171
.7	1.03211	1.0387	1.0394	1.0395
.8	1.06613	1.0882	1.0921	1.0927
.9	1.13728	1.2198	1.2458	1.2528
		Q = .8		
.5	1.00757	1.0075	1.0078	1.0078
.6	1.01772	1.0188	1.0188	1.0188
.7	1.03866	1.0424	1.0426	1.0427
.8	1.08269	1.0966	1.0979	1.0980
.9	1.18491	1.2477	1.2596	1.2617
		Q = .9		
.5	1.00581	1.0058	1.0058	1.0058
.6	1.01384	1.0140	1.0140	1.0140
.7	1.03107	1.0319	1.0319	1.0319
.8	1.07000	1.0736	1.0737	1.0737
.9	1.17573	1.1980	1.1998	1.1999

Table 6.4. Comparison of ${}_2\overline{Y}_{\alpha_i}$ with ${}_7\overline{Y}_{\alpha_i}$

ρ	M_1/C_1	M_2/C_2	M_3/C_3	M_4/C_4
		Q = .5		
.5	1.07142	1.0051	1.003	1.0000
.6	1.05134	1.0057	1.0006	1.0001
.7	1.03311	1.0056	1.0009	1.0002
.8	1.01729	1.0044	1.0011	1.0003
.9	1.00525	1.0021	1.0009	1.0003
		Q = .6		
.5	1.07058	1.0444	1.0433	1.0432
.6	1.05289	1.0348	1.0335	1.0334
.7	1.03574	1.0256	1.0245	1.0244
.8	1.01987	1.0164	1.0160	1.0159
.9	1.00655	1.0069	1.0074	1.0077
		Q = .7		
.5	1.06363	1.0551	1.0549	1.0549
.6	1.04979	1.0442	1.0440	1.0440
.7	1.03578	1.0335	1.0335	1.0335
.8	1.02155	1.0225	1.0228	1.0229
.9	1.00797	1.0104	1.0115	1.0118
		Q = .8		
.5	1.04999	1.0480	1.0480	1.0480
.6	1.04122	1.0402	1.0402	1.0402
.7	1.03183	1.0320	1.0321	1.0321
.8	1.02129	1.0231	1.0233	1.0234
.9	1.00931	1.0122	1.0131	1.0132
		Q = .9		
.5	1.02903	1.0288	1.0288	1.0288
.6	1.02541	1.0253	1.0253	1.0253
.7	1.02143	1.0216	1.0217	1.0217
.8	1.01652	1.0173	1.0173	1.0173
.9	1.00943	1.0110	1.0112	1.0113

occasions, increase. It also increases as Q increases, but decreases when Q approaches 1. Therefore, the symmetric estimator is good when Q is near .5 and ρ is close to .5.

From Table 6.4, for every i, the complete estimator is better than the modified composite estimator over the values of ρ and Q under consideration. The ratio decreases as ρ increases. When Q is .5 or .6, the ratio decreases as the number of previous occasions increases. However, when $Q > .7$ $\rho > .7$, the ratio increases as the number of previous occasions increases. This means that the modified composite estimator is good when ρ is high, and when the number of previous occasions is large.

In conclusion, the complete estimator is the best among the three estimators for all i .

Finally, it may be useful to observe the number of previous occasions required in the estimators ${}_2\bar{Y}_{\alpha_i}$ or ${}_5\bar{Y}_{\alpha_i}$ such that the estimators have the same precision as ${}_7\bar{Y}_{\alpha_3}$. Table 6.5 provides the values of S_i, M_i, C_i for this purpose.

From the table, the complete estimator using only one previous occasion is even more efficient than the symmetric estimator using seven previous occasions when $Q \geq 6$. Also M_4 is larger than C_2 when $Q \geq .6$ except for the points of ρ and Q marked * . This implies that the complete estimator using two previous occasions is more efficient than the modified composite estimator using more than two previous occasions.

In the second part, the complete estimator will be compared with the modified composite estimator for all values of Q, $.2 \leq Q \leq 1$, when $i = 2$. Since the preferred number of previous occasions for the complete

Table 6.5. Values of Q_i, M_i and C_i

ρ	S_1	S_2	S_3	S_4	. . .	S_7
			Q = .5			
.5	.2333	.2321	.2320	.2320		.2320
.6	.2252	.2225	.2222	.2222		.2222
.7	.2150	.2094	.2084	.2083		.2083
.8	.2023	.1911	.1884	.1877		.1875
.9	.1865	.1649	.1568	.1537		.1519
			Q = .6		. . .	
.5	.2249	.2243	.2243	.2243		.2243
.6	.2179	.2165	.2164	.2164		.2264
.7	.2093	.2065	.2062	.2062		.2062
.8	.1989	.1936	.1929	.1928		.1928
.9	.1863	.1770	.1753	.750		.1749
			Q = .7			
.5	.1987	.1984	.1984	.1984	. . .	.1984
.6	.1935	.1930	.1930	.1930		.1930
.7	.1873	.1863	.1862	.1862		.1862
.8	.1800	.1781	.1779	.1779		.1779
.9	.1714	.1681	.1678	.1778		.1678
			Q = .8			
.5	.1535	.1534	.1534	.1534	. . .	.1534
.6	.1506	.1505	.1505	.1505		.1505
.7	.1472	.1469	.1469	.1469		.1469
.8	.1431	.1427	.1427	.1427		.1427
.9	.1384	.1378	.1378	.1378		.1378
			Q = .9			
.5	.0879	.0879	.0879	.0879	. . .	.0879
.6	.0870	.0870	.0870	.0870		.0870
.7	.0860	.0859	.0859	.0859		.0859
.8	.0847	.0847	.0847	.0847		.0847
.9	.0833	.0833	.0833	.0833		.0833

(Table 6.5 continued)

(Table 6.5 continued)

ρ	Q = .5 M₂				Q = .5		
ρ	M_2	M_3	M_4		C_1	C_2	C_3
.5	.2333	.2321	.2320*		.2333	.2321	.2320
.6	.2238	.2224*	.2222*		.2252	.2225	.2222
.7	.2106	.2086*	.2083*		.2150	.2094	.2084
.8	.1920	.1886*	.1877*		.2023	.1911	.1884
.9	.1652	.1570*	.1538		.1865	.1649	.1568
	Q = .6				Q = .6		
.5	.2333	.2330	.2330		.2241	.2234	.2233
.6	.2218	.2213	.2213		.2161	.2143	.2142
.7	.2068	.2061	.2060		.2057	.2016	.2011
.8	.1864	.1847	.1844		.1921	.1834	.1818
.9	.1568	.1511	.1493*		.1741	.1557	.1500
	Q = .7				Q = .7		
.5	.2079	.2078	.2078		.1974	.1970	.1970
.6	.1981	.1981	.1981		.1907	.1897	.1897
.7	.1853	.1851	.1851		.1815	.1793	.1791
.8	.1673	.1666	.1666		.1688	.1636	.1629
.9	.1393	.1363	.1355		.1507	.1378	.1347
	Q = .8				Q = .8		
.5	.1595	.1595	.1595		.1523	.1522	.1522
.6	.1536	.1536	.1536		.1480	.1477	.1477
.7	.1454	.1454	.1454		.1417	.1409	.1409
.8	.1331	.1330	.1330		.1322	.1301	.1299
.9	.1118	.1108	.1106		.1169	.1104	.1093
	Q = .9				Q = .9		
.5	.0899	.0899	.0899		.0874	.0874	.0874
.6	.0880	.0880	.0880		.0858	.0858	.0858
.7	.0851	.0851	.0851		.0834	.0833	.0833
.8	.0802	.0802	.0802		.0792	.0789	.0789
.9	.0703	.0703	.0703		.0709	.0695	.0694

estimator is two, for all values of Q, it is sufficient to consider only the comparison of the estimators when $i = 2$.

Table 6.6 gives the ratio of the variance of ${}_2\bar{Y}_{\alpha_2}$ over the variance of ${}_7\bar{Y}_{\alpha_2}$.

Table 6.6. Comparison of ${}_7\bar{Y}_{\alpha_2}$ with ${}_2\bar{Y}_{\alpha_2}$ when $Q \geq .2$

$$V({}_2\bar{Y}_{\alpha_2})/V({}_7\bar{Y}_{\alpha_2}) = M_2/C_2$$

ρ \ Q	.2	.3	.4	.5	.6	.7	.8	.9
.5	1.0947	1.0940	1.0647	1.0051	1.0444	1.0551	1.0480	1.0288
.6	1.0691	1.0706	1.0503	1.0057	1.0348	1.0442	1.0402	1.0253
.7	1.0450	1.0477	1.0356	1.0056	1.0256	1.0335	1.0320	1.0216
.8	1.0235	1.0263	1.0209	1.0044	1.0164	1.0225	1.0231	1.0173
.9	1.0070	1.0085	1.0074	1.0021	1.0069	1.0104	1.0122	1.0110

The ratio is larger when ρ is small and Q is away from .5. This implies that when $i = 2$ the complete estimator is uniformly better than the modified composite estimator for all values of Q and ρ. The same conclusion can be observed by comparing Table 5.5 and Table 6.2.

In summary, the complete estimator is chosen as the best estimator among the three estimators. The modified estimator is generally better than the symmetric estimator, even though the symmetric estimator is better than the modified composite estimator for some region of ρ and Q.

6.4. A Complete Estimator in Multi-stage Sampling

The theory of the complete estimator in multi-stage sampling will be developed for the estimator of $\bar{Y}_\alpha$ defined in (6.11), using two previous occasions.

$$ {}_7\hat{\hat{Y}}_{\alpha_2} = a_o\hat{\hat{Y}}_{\alpha 1} + b_o\hat{\hat{Y}}_{\alpha 2} + a_1\hat{\hat{Y}}_{(\alpha-1)1} + b_1\bar{\bar{Y}}_{(\alpha-1)2} + a_2\hat{\hat{Y}}_{(\alpha-2)1} + b_2\hat{\hat{Y}}_{(\alpha-2)2} \tag{6.11} $$

where, $a_o + b_o = 1 \qquad a_i + b_i = 0 \qquad i = 1, 2$

$\hat{\hat{Y}}_{(\alpha-i)j}$ is defined in Lemma 6.

By Lemma 6, ${}_7\bar{\bar{Y}}_{\alpha_2}$ is an unbiased estimator of $\bar{\bar{Y}}_{\alpha}$, i.e.,

$$ E\ {}_7\hat{\hat{Y}}_{\alpha_2} = \bar{\bar{Y}}_{\alpha} \quad . $$

In order to derive the variance of ${}_7\hat{\hat{Y}}_{\alpha_2}$, let $\alpha = 3$. Then

$$ {}_7\hat{\hat{Y}}_{3.} = \hat{\hat{Y}}_{32} + a_o\ (\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}) + a_1(\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) + a_2(\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) $$

$$ \begin{aligned} V({}_7\hat{\hat{Y}}_{3.}) = {} & V(\hat{\hat{Y}}_{32}) + a_o^2\ V\ (\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}) + a_1^2 V(\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) + a_2^2 V(\hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) \\ & + 2a_o\ \mathrm{cov}\ (\hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}) + 2a_1 \mathrm{cov}\ (\hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) \\ & + 2a_2\ \mathrm{cov}\ (\hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) + 2a_o a_1\ \mathrm{cov}\ (\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) \\ & + 2a_o a_2\ \mathrm{cov}\ (\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}) + 2a_1 a_2 \mathrm{cov}(\hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}, \hat{\hat{Y}}_{11} - \hat{\hat{Y}}_{12}). \end{aligned} $$

By analogy with 4.4., and by Lemma 7 and 8,

$$ \mathrm{cov}\ (\hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}) = -\frac{N}{N-1}\ \frac{\sigma^2_{3.1}}{Qn} - \frac{\sigma^2_{3.2}}{Qn} = -\frac{N}{N-1}\ \frac{\bar{\sigma}^2_3}{Qn} $$

$$ \begin{aligned} V(\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}) &= \frac{\sigma^2_{3.1}}{Pn}\ \left(\frac{N-Pn}{Pn}\right) + \frac{\sigma^2_{3.2}}{Pn} + \frac{\sigma^2_{3.1}}{Qn}\ \left(\frac{N-Qn}{N-1}\right) + \frac{\sigma^2_{3.2}}{Qn} + 2\ \frac{\sigma^2_{3.1}}{N-1} \\ &= \frac{N}{N-1}\ \frac{\bar{\sigma}^2_3}{nPQ} \end{aligned} $$

$$ \mathrm{cov}\ (\hat{\hat{Y}}_{31} - \hat{\hat{Y}}_{32}, \hat{\hat{Y}}_{21} - \hat{\hat{Y}}_{22}) = \frac{N}{N-1}\ \frac{\sigma_{23.1}}{Pn} + \frac{\sigma_{23.2}}{Pn} = \frac{N}{N-1}\ \frac{\bar{\sigma}_{23}}{Pn} \quad . $$

Similarly, $\text{cov}\,(\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{11} - \bar{Y}_{12}) = \frac{N}{N-1} \frac{\bar{\sigma}_{13}}{Pn}$

where, $\bar{\sigma}^2_{3-i}$ $\sigma_{(3-i).1}$ $\sigma_{(3-i)(3-j),1}$ are defined as in 4.4 and Lemma 7 and 8. Hence,

$$V({}_7\hat{\bar{Y}}_{3.}) = \frac{\sigma^2_{3.1}}{Qn}\left(\frac{N-Qn}{N-1}\right) + \frac{\sigma^2_{3.2}}{Qn} + a_o^2 \frac{N}{N-1} \frac{\bar{\sigma}^2_3}{nPQ} + a_1^2 \frac{N}{N-1} \frac{\bar{\sigma}^2_2}{nPQ}$$

$$+ a_2^2 \frac{N}{N-1} \frac{\bar{\sigma}^2_1}{nPQ} - 2a_o \frac{N}{N-1} \frac{\bar{\sigma}^2_3}{Qn} + 2a_o a_1 \frac{N}{N-1} \frac{\bar{\sigma}_{23}}{Pn}$$

$$+ 2a_o a_2 \frac{N}{N-1} \frac{\bar{\sigma}_{13}}{Pn} + 2\, a_1 a_2 \frac{N}{N-1} \frac{\bar{\sigma}_{12}}{Pn} \quad .$$

From (6.4), the optimum value of a_i is as follows.

$$a_{2w} = \frac{A_{31}}{|A_3|} P\bar{\sigma}^2_3$$

$$a_{1w} = \frac{A_{32}}{|A_3|} P\bar{\sigma}^2_3$$

$$a_{ow} = \frac{A_{33}}{|A_3|} P\bar{\sigma}^2_3$$

where $|A_3|$ and the cofactors are defined in (6.4). Hence, the variance of the optimum estimator ${}_7\bar{Y}_{3.w}$ in multi-stage sampling is

$$V({}_7\hat{\bar{Y}}_{3.w}) = \frac{\sigma^2_{3.1}}{Qn}\left(\frac{N-Qn}{N-1}\right) + \frac{\sigma^2_{3.2}}{Qn} - \frac{N}{N-1} \frac{\bar{\sigma}^4_3}{nPQ} \frac{P^2}{|A_3|^2} \left(A_{33}{}^2 \right.$$

$$\left. - \bar{\sigma}^2_2 A_{32}{}^2 - \bar{\sigma}^2_1 A_{31}{}^2 - 2Q\, \bar{\sigma}_{12} A_{31} \cdot A_{32} \right) .$$

Therefore

$$V({}_7\hat{\bar{Y}}_{\alpha_2 w}) = \frac{\sigma^2_{\alpha.1}}{Qn}\left(\frac{N-Qn}{N-1}\right) + \frac{\sigma^2_{\alpha.2}}{Qn} - \frac{N}{N-1} \frac{\bar{\sigma}^2_\alpha}{nPQ} \frac{P^2}{|A_3|^2} \left(A_{32}{}^2 \right.$$

$$- \bar{\sigma}^2_{\alpha-1} A_{32}{}^2 - \bar{\sigma}^2_{\alpha-2} A_{31}{}^2 - 2Q\bar{\sigma}_{\alpha-1,\alpha-2} A_{31} \cdot A_{32})$$

where the cofactors are defined appropriately.

7. SUMMARY AND CONCLUSIONS

7.1. Summary

In this thesis, estimation theory for samples selected on successive occasions is treated with multi-stage sampling without replacement of units at each draw. Two designs were studied: (1) the proposed design which is a traditional scheme and (2) a new design. Theory was developed for the two different designs as follows:

1. Since, for a population changing over time, variances and covariances between variates of subsets (specified by the design) on any two different occasions are fundamental quantities, several lemmas for obtaining these variances and covariances were derived.

2. Estimation theory for samples selected on two successive occasions provides a basis for the theory applicable to more than two occasions. Accordingly, the properties of four classes of estimators and comparisons among them have been discussed for the two successive occasions case. It was found that the percent gain of each estimator over the simple estimator of $\bar{\bar{Y}}_\alpha$ increased as ρ increased, and the gain always attained a maximum when $Q > .5$. In other words, the optimum value of Q for each estimator was always larger than .5.

Comparing these four estimators, the general linear estimator was uniformly the best when $0 < Q < 1$, $\rho > .5$, and the modified composite estimator was the next best when $0 < Q < 1$, $\rho > .7$.

In addition, the symmetric estimator has been discussed for the two successive occasions case. Next to the general linear estimator, the symmetric estimator and the modified composite estimator were found to be more efficient than the other estimators studied. The symmetric estimator is preferred when $\rho < .75$ and $.5 < Q < .7$, the modified composite estimator is preferred when $\rho > .75$ and $Q > .7$.

3. When sampling on more than two occasions with the proposed design, the optimum form and properties of the general linear estimator are difficult to determine. Hence, the symmetric estimator and the modified composite estimator have been discussed for sampling on more than two occasions.

The percent gain of each estimator over the sample mean for a single occasion increased as ρ and i increased. Also the optimum value of Q was close to .5. The preferred number of previous occasions i was determined to be two for both estimators.

The symmetric estimator is preferred when Q and ρ are small and the modified composite estimator is preferred when ρ is large. However, the modified composite estimator is generally better than the symmetric estimator when i is large.

4. It has been seen that neither the symmetric estimator nor the modified composite estimator is uniformly better than the other. The properties of the complete estimator have been investigated for the new design for values of $\rho > .5$ and $0 < Q < 1$. The preferred number of previous occasions was again two. The complete estimator was found to be definitely better than the previous two estimators (proposed design) for any given value of i when $Q \geq 5$. When $i = 2$, the complete estimator is uniformly better than the modified composite estimator for all values of Q and $\rho \geq .5$. Moreover, the complete estimator, using only the

information available from one previous occasion, is preferred to the symmetric estimator using data from two or more previous occasions and the complete estimator using two previous occasions is preferred to the modified composite estimator using data from three or more previous occasions.

7.2. Conclusions

Estimation theory is concerned with choosing an estimator which has a smaller variance. From the results derived in this thesis, it has been shown that, for sampling on two successive occasions, the general linear estimator as defined herein and based on the proposed design has a smaller variance than the other estimators studied for the same design. For sampling on more than two occasions, the complete estimator based on the new design has a smaller variance than the symmetric estimator or the modified composite estimator based on the proposed design.

This indicates, as suspected in this thesis since the complete estimator is the general linear estimator (defined in Chapter 5), that the complete estimator has a smaller variance than any other linear unbiased estimator. This result, if valid, coincides with the theorem given by Eckler [3] for sampling with replacement.

With respect to the effects of the sampling design, the new design has the following advantages:

(1) when the complete estimator (as the general linear estimator) is incorporated with the new design, the number of the coefficients to be determined, in order to find the optimum estimator, is small.

(2) the variance formula of the complete estimator derived in this thesis is valid for any value of Q. However, the variance formula of

the general linear estimator under the proposed design changes when the discarded fraction Q varies.

7.3. Recommendation for Future Research

Regarding the new design studied in this thesis, two points are noted:

(1) the first-stage units are selected with equal probabilities without replacement,

(2) the second-stage units within the matched first-stage units are retained in the sample for all successive occasions.

However, the following design for sampling on successive occasions is more general and is suggested as worthy of future research:

(1) a partial retention scheme for the first-stage units is carried out in accordance with the new design, where the first-stage units are selected with probability proportional to size, and

(2) a partial retention scheme such as used in the proposed design in this thesis or the rotation scheme in the C.P.S. redesign is applied to the second-stage units within the matched first-stage units under the new design.

It is considered advisable to make a study of the properties and the efficiency of the complete estimator in comparison with alternative estimators for the suggested design. Some comparisons of the efficiency of the suggested design also should be made with alternative sampling designs.

8. LIST OF REFERENCES

1. Cochran, W. G. 1963. Sampling Techniques, Second Edition. John Wiley and Sons, New York.

2. Des Raj. 1965. On sampling over two occasions with proportionate to size. Annals of Mathematical Statistics, 36:327-330.

3. Eckler, A. R. 1955. Rotation sampling. Annals of Mathematical Statistics, 36:664-685.

4. Hansen, M. H., W. N. Hurwitz, H. Nisselson and J. Steinberg. 1955. The redesign of the census current population survey. Journal of the American Statistical Association, 50:701-719.

5. Hansen, M.H., W. H. Hurwitz, and W. G. Madow. 1953. Sample Survey Methods and Theory, Vol. I and II. John Wiley and Sons, New York.

6. Horvitz, D. G. and D. J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47:663-685.

7. Jessen, R. J. 1942. Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station Research Bulletin, No. 304. Ames, Iowa.

8. Koop, J. C. 1963. On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite universes. Metrica 7(2 and 3):81-114 and 165-204.

9. Onate, B. T. 1960. Development of multi-stage designs for statistical surveys in the Phillipines. Mimeo-Multilith Series, No. 3, Statistical Laboratory, Iowa State University, Ames, Iowa

10. Patterson, H. D. 1950. Sampling on successive occasions with partial replacement of units. Journal of the Royal Statistical Society, Series B, 12:241-255.

11. Purakam, Niyom. 1966. Multi-stage sampling on successive occasions where first-stage units are drawn with unequal probabilities and with replacement. Unpublished Ph.D. thesis, Department of Experimental Statistics, North Carolina State University. Raleigh, North Carolina.

12. Rao, J.N.K. and J. E. Graham. 1964. Rotation designs for sampling on repeated occasions. Journal of the American Statistical Association, 59:492-509.

13. Tikkiwal, B. D. 1955. Multiphase sampling on successive occasions. Unpublished Ph.D. thesis, Department of Experimental Statistics, North Carolina State University, Raleigh, North Carolina.

14. Yates, A. 1960. Sampling Method for Census and Surveys. Charles Griffin and Co., London.

9. APPENDICES

9.1. Estimation of a_w, c_w and $V({}_1\bar{Y}_w)$

9.1.1. Estimation of a_w

The optimum weight of a

$$a_w = \frac{PQ\rho}{1-Q^2r^2} \; \frac{\sigma_2}{\sigma_1}$$

is estimated with small bias by

$$\hat{a}_w = \frac{PQr}{1-Q^2r^2} \; \frac{s_2}{s_1} \quad .$$

Proof:

$\frac{r}{1-Q^2r^2}$ can be expressed as follows:

$$\frac{r}{1-Q^2r^2} = (\rho+r-\rho)\,[1-Q^2\rho^2 + 1 - Q^2r^2 - (1-Q^2\rho^2)]^{-1}$$

$$= \frac{\rho}{1-Q^2\rho^2}\;(1 + \frac{r-\rho}{\rho})\left[1 - \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2}\right]^{-1} \quad .$$

Since $\frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2} < 1$,

$$\frac{r}{1-Q^2r^2} = \frac{\rho}{1-Q^2\rho^2}\,(1 + \frac{r-\rho}{\rho})(1 + \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2} + \frac{Q^4(r^2-\rho^2)^2}{(1-Q^2\rho^2)^2} + \ldots) \quad .$$

Also

$$\frac{s_2}{s_1} = \frac{\sigma_2}{\sigma_1}\,(1 + \frac{s_2-\sigma_2}{\sigma_2} - \frac{s_1-\sigma_1}{\sigma_1} - \frac{(s_1-\sigma_1)(s_2-\sigma_2)}{\sigma_1\sigma_2} + \frac{(s_1-\sigma_1)^2}{\sigma_1^2} + \ldots).$$

Therefore,

$$PQ \frac{r}{1-Q^2r^2} \frac{s_2}{s_1} \doteq PQ \frac{\rho}{1-Q^2\rho^2} \frac{\sigma_2}{\sigma_1} \Big(1 + \frac{r-\rho}{\rho} + \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2}$$

$$+ \frac{s_2-\sigma_2}{\sigma_2} + \frac{(r-\rho)(s_2-\sigma_2)}{\rho \quad \sigma_2} + \frac{Q^2(r^2-\rho^2)(s_2-\sigma_2)}{(1-Q^2\rho^2) \quad \sigma_2}$$

$$- \frac{s_1-\sigma_1}{\sigma_1} - \frac{(r-\rho)}{\rho}\frac{(s_1-\sigma_1)}{\sigma_1} - \frac{Q^2(r^2-\rho^2)(s_1-\sigma_1)}{(1-Q^2\rho^2)\sigma_1}$$

$$+ \frac{(s_1-\sigma_1)^2}{\sigma_1^2} + \frac{(r-\rho)(s_1-\sigma_1)^2}{\rho \quad \sigma_1^2} + \frac{Q^2(r^2-\rho^2)}{(1-Q^2\rho^2)}\frac{(s_1-\sigma_1)^2}{\sigma_1^2} + \dots \Big) .$$

Since, under the assumption of normality,

$$E(s) = \sigma + O(\tfrac{1}{n}) = \sigma - \frac{\sigma}{2n} - \frac{\sigma^2}{4n^2}$$

$$E(r) = \rho(1 - \frac{1-\rho^2}{2n} + O\ (\frac{1}{n^2})) = \rho - \frac{\rho(1-\rho^2)}{2n} + O(\frac{1}{n^2})$$

$$E(r^2) = \rho^2 + \frac{(1-\rho^2)^2}{n}$$

to the order $\frac{1}{n^2}$, the expectation is

$$E(PQ \frac{r}{1-Q^2r^2} \frac{s_2}{s_1}) = PQ \frac{\rho}{1-Q^2\rho^2} \frac{\sigma_2}{\sigma_1} + E(R) + O(\frac{1}{n^2})$$

with bias = $E(R)$,

where $R = \frac{r-\rho}{\rho} + \frac{Q^2(r^2-\sigma^2)}{1-Q^2\rho^2} + \frac{s_2-\sigma_2}{\sigma_2} - \frac{s_1-\sigma_1}{\sigma_1} + \frac{(s_1-\sigma_1)^2}{\sigma_1^2} - \frac{(s_1-\sigma_1)(s_2-\sigma_2)}{\sigma_1\sigma_2}$.

9.1.2. Estimation of C_w

The optimum weight of C

$$C_w = \frac{P}{1-Q^2\rho^2}$$

is estimated with small bias by

$$\hat{C}_w = \frac{P}{1-Q^2r^2} \quad .$$

Proof: As in 9.1.1,

$$\frac{1}{1-Q^2r^2} = \frac{1}{1-Q^2\rho^2}\left(1 + \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2} + \frac{Q^4(r^2-\rho^2)^2}{(1-Q^2\rho^2)^2} + \ldots\right).$$

Hence,

$$\frac{P}{1-Q^2r^2} = \frac{P}{1-Q^2\rho^2} \quad \text{with the bias} = \frac{Q^2(1-\rho^2)^2}{(1-Q^2\rho^2)n}$$

to the order $\frac{1}{n^2}$.

9.1.3. Estimation of $V({}_1\bar{Y}_w)$

The variance of ${}_1\bar{Y}_w$

$$V({}_1\bar{Y}_w) = \frac{N}{N-1}\,\frac{\sigma_2^2}{n}\left(\frac{1-Q\rho^2}{1-Q^2\rho^2}\right) - \frac{\sigma_2^2}{N-1} \quad \text{is estimated approximately}$$

by

$$\hat{V}({}_1\bar{Y}_w) = \frac{s_2^2}{n}\left(\frac{1-Qr^2}{1-Q^2r^2}\right) - \frac{s_2^2}{N} \; .$$

Proof:

$\frac{1-Qr^2}{1-Q^2r^2}$ can be expressed as follows:

$$\frac{1-Qr^2}{1-Q^2r^2} = \frac{1-Q\rho^2}{1-Q^2\rho^2}\left(1 - \frac{Q(r^2-\rho^2)}{1-Q\rho^2} + \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2} - \frac{Q^3(r^2-\rho^2)^2}{(1-Q\rho^2)(1-Q^2\rho^2)} + \ldots\right).$$

Since the bias is small for large n,

$$\hat{V}\,({}_1\bar{\bar{Y}}_w) = \frac{s_2^2}{n}\left(\frac{1-Qr^2}{1-Q^2r^2}\right) - \frac{s_2^2}{N}$$

with bias $= \dfrac{\sigma_2^2}{n}\left(\dfrac{1-Q\rho^2}{1-Q^2\rho^2}\right)\dfrac{Q(1-\rho^2)^2}{n}\left(\dfrac{Q}{1-Q^2\rho^2} - \dfrac{1}{1-Q\rho^2}\right)$

to the order $\dfrac{1}{n^3}$.

9.1.4. Variance of ${}_1\tilde{\bar{\bar{Y}}}_w$

$$V({}_1\tilde{\bar{\bar{Y}}}_w) \doteq V({}_1\bar{\bar{Y}}_w)$$

Proof: In 4.2.1, ${}_1\tilde{\bar{\bar{Y}}}_w$ was defined as

$${}_1\tilde{\bar{\bar{Y}}}_w = \hat{a}_w\,(\bar{Y}_{11} - \bar{Y}_{12}) + \hat{c}_w(\bar{Y}_{21} - \bar{Y}_{22}) + \bar{Y}_{22} \quad .$$

By substituting $\hat{a}_w$ and $\hat{c}_w$, ${}_1\tilde{\bar{\bar{Y}}}_w$ can be written as follows.

$${}_1\tilde{\bar{\bar{Y}}}_w = PQ\,\frac{\rho}{1-Q^2\rho^2}\;\frac{\sigma_2}{\sigma_1}\left(1 + \frac{s_2-\sigma_2}{\sigma_2} - \frac{s_1-\sigma_1}{\sigma_1} + \frac{r-\rho}{\rho} + \ldots\right)(\bar{Y}_{11} - \bar{Y}_{12})$$

$$+ \;\frac{P}{1-Q^2\rho^2}\left(1 + \frac{Q^2(r^2-\rho^2)}{1-Q^2\rho^2} + \ldots\right)(\bar{Y}_{21} - \bar{Y}_{22}) + \bar{Y}_{22}$$

$$= {}_1\bar{\bar{Y}}_w + (\text{Remainder R}) + O\left(\frac{1}{n^2}\right) \quad ,$$

where the remainder R is

$$R = PQ\,\frac{\rho}{1-Q^2\rho^2}\;\frac{\sigma_2}{\sigma_1}\left(\frac{s_2-\sigma_2}{\sigma_2} - \frac{s_1-\sigma_1}{\sigma_1} + \frac{r-\rho}{\rho}\right)(\bar{Y}_{11} - \bar{Y}_{12})$$

$$+ \;\frac{PQ^2(r^2-\rho^2)}{(1-Q^2\rho^2)^2}\;(\bar{Y}_{21} - \bar{Y}_{22}) \quad .$$

Then the variance of the remainder R is

$$V(R) = \frac{P^2Q^2\rho^2}{(1-Q^2\rho^2)^2} \quad \frac{\sigma_2^2}{\sigma_1^2}\Bigg[\frac{V(s_2-\sigma_s)(\overline{Y}_{11}-\overline{Y}_{12})}{\sigma_2^2} - \frac{V(s_1-\sigma_1)(\overline{Y}_{11}-\overline{Y}_{12})}{\sigma_1^2}$$

$$+ \frac{V(r-\rho)(\overline{Y}_{11}-\overline{Y}_{12})}{\rho^2} - 2\,\frac{\text{cov}[s_2-\sigma_2)(\overline{Y}_{11}-\overline{Y}_{12}),(s_1-\sigma_1)(\overline{Y}_{11}-\overline{Y}_{12})]}{\sigma_1\ \sigma_2}$$

$$+ 2\,\frac{\text{cov}\ [s_2-\sigma_2)(\overline{Y}_{11}-\overline{Y}_{12}),(r-\rho)(\overline{Y}_{11}-\overline{Y}_{12})]}{\rho\ \sigma_2}$$

$$- 2\,\frac{\text{cov}[(s_1-\sigma_1)(\overline{Y}_{11}-\overline{Y}_{12}),\ (r-\rho)(\overline{Y}_{11}-\overline{Y}_{12})]}{\rho\ \sigma_1}$$

$$+ \frac{P^2Q^4}{(1-Q^2\rho^2)^4}\ V\,[(r^2-\rho^2)(\overline{Y}_{11}-\overline{Y}_{22})]$$

$$+ 2\,\frac{P^2Q^3\rho}{(1-Q^2\rho^2)^3}\ \frac{\sigma_2}{\sigma_1}\Bigg[\ \text{cov}\ \frac{(s_2-\sigma_2)(\overline{Y}_{11}-\overline{Y}_{12}),(r^2-\rho^2)(\overline{Y}_{21}-\overline{Y}_{22})}{\sigma_2}$$

$$- \frac{\text{cov}\ (s_1-\sigma_1)(\overline{Y}_{11}-\overline{Y}_{12}),(r^2-\rho^2)(\overline{Y}_{21}-\overline{Y}_{22})}{\sigma_1}$$

$$+ \frac{\text{cov}(r-\rho)(\overline{Y}_{11}-\overline{Y}_{12}),(r^2-\rho^2)(\overline{Y}_{21}-\overline{Y}_{22})}{\rho}\ \Bigg] \quad .$$

Consider the variances and covariances in V(R) for large n.

1. $V(s_2-\sigma_2)(\overline{Y}_{11}-\overline{Y}_{12})$

Let $u = s_2 - \sigma_2$

$w = \overline{Y}_{11} - \overline{Y}_{12}$,

then by Taylor's theorem

$$u \cdot w = Eu \cdot Ew + (u - Eu)Ew + (w - Ew)Eu .$$

Since $Ew = 0$, $u.w = w(Eu)$

and $V(u.w) = V(w)(Eu)^2$.

Now,

$$Eu = E(s_2 - \sigma_2) = 0(\frac{1}{n})$$

$$V(u.w) = \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} \cdot 0(\frac{1}{n}) \text{ , which is negligible.}$$

2. $V(r-\rho)(\overline{Y}_{11} - \overline{Y}_{12})$

Similarly, let $u = r-\rho$

$w = \overline{Y}_{11} - \overline{Y}_{12}$.

Since $Eu = E(r-\rho) = \frac{\rho(1-\rho^2)}{2n} + 0\ (\frac{1}{n^2})$,

$$V(r-\rho)(\overline{Y}_{11} - \overline{Y}_{12}) = \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} (\frac{1-\rho^2}{2n})^2 \text{ , which is negligible.}$$

3. Covariance of $(s_2-\sigma_2)(\overline{Y}_{11} - \overline{Y}_{12})$, $(r-\rho)(\overline{Y}_{11} - \overline{Y}_{12})$

Let $u = s_2 - \sigma_2$

$w = \overline{Y}_{11} - \overline{Y}_{12}$

$x = r-\rho$,

then , $u \cdot w = w(Eu) = (\overline{Y}_{11} - \overline{Y}_{12}) \cdot 0\ (\frac{1}{n})$

$$x \cdot w = w(Ex) = (\overline{Y}_{11} - \overline{Y}_{12}) \frac{\rho(1-\rho^2)}{2n} .$$

Hence, $\text{cov}\ (u.w\ ,\ x.w) = (Eu)(Ex)\ V(x) = 0(\frac{1}{n}) \frac{\rho(1-\rho^2)}{2n} \frac{N}{N-1} \frac{\sigma_1^2}{nPQ}$,

which is also negligible.

4. $V(r^2-\rho^2)(\bar{Y}_{21}-\bar{Y}_{22})$

Let $$u = r^2 - \rho^2$$
$$w = \bar{Y}_{21} - \bar{Y}_{22} ,$$

then

$$\therefore u.w = w(Eu) = (\bar{Y}_{21} - \bar{Y}_{22}) \frac{(1-\rho^2)^2}{n}$$

$$V(u.w) = \frac{(1-\rho^2)^4}{n^2} \frac{N}{N-1} \frac{\sigma_2^2}{nPQ} \quad , \quad \text{which is negligible.}$$

Similarly, the others also can be proved to be negligible.

Hence, for large n ,

$$V({}_1\tilde{\bar{Y}}_w) \doteq V({}_1\bar{Y}_w) \quad .$$

9.2. Bias of ${}_4\bar{Y}_2$

$${}_4\bar{Y}_2 = a(\bar{Y}_1 \cdot \frac{\bar{Y}_{21}}{\bar{Y}_{12}}) + b\,\bar{Y}_2 \qquad \text{where } a + b = 1$$

$$= a(P\bar{Y}_{12} + Q\bar{Y}_{11}) \frac{\bar{Y}_{21}}{\bar{Y}_{12}} + b(P\bar{Y}_{21} + Q\bar{Y}_{22})$$

$$= aQ\,(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}}) + P\bar{Y}_{21} + bQ\bar{Y}_{22} \qquad (9.1)$$

$$E_4\bar{Y}_2 = aQ\,E\,\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}} + (P+bQ)\,\bar{\bar{Y}}_2 \qquad \text{(From Lemma 1)} .$$

Now consider $E\,(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}})$,

which can be written as follows:

$$\frac{\bar{Y}_{11}\cdot\bar{Y}_{21}}{\bar{Y}_{12}} = \frac{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}{\bar{\bar{Y}}_1}(1+\Delta\bar{Y}_{11})(1+\Delta\bar{Y}_{21})\frac{1}{(1+\Delta\bar{Y}_{12})}$$

where

$$\Delta\bar{Y}_{ij} = \frac{\bar{Y}_{ij}-\bar{\bar{Y}}_i}{\bar{\bar{Y}}_i}$$

$$= \bar{\bar{Y}}_2(1+\Delta\bar{Y}_{11}+\Delta\bar{Y}_{21}+\Delta\bar{Y}_{11}\Delta\bar{Y}_{21})(1-\Delta\bar{Y}_{12}+\frac{(\Delta\bar{Y}_{12})^2}{1+\Delta\bar{Y}_{12}}) .$$

Since $1+\Delta\bar{Y}_{11} = \frac{\bar{Y}_{11}}{\bar{\bar{Y}}_1}$, $\frac{1}{1+\Delta\bar{Y}_{12}} = \frac{\bar{\bar{Y}}_1}{\bar{Y}_{12}}$

$$\frac{\bar{Y}_{11}\cdot\bar{Y}_{21}}{\bar{Y}_{12}} = \bar{\bar{Y}}_2(1+\Delta\bar{Y}_{11}+\Delta\bar{Y}_{21}+\Delta\bar{Y}_{11}\Delta\bar{Y}_{21})(1-\Delta\bar{Y}_{12}) + \frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\frac{(\bar{Y}_{12}-\bar{\bar{Y}}_1)^2}{\bar{\bar{Y}}_1^{\,2}} \qquad (9.2)$$

$$\therefore\ E\left(\frac{\bar{Y}_{11}\cdot\bar{Y}_{21}}{\bar{Y}_{12}}\right) = \bar{\bar{Y}}_2\left(1+\frac{\operatorname{cov}(\bar{Y}_{11}\bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}-\frac{\operatorname{cov}(\bar{Y}_{11}\bar{Y}_{12})}{\bar{\bar{Y}}_1^2}-\frac{\operatorname{cov}(\bar{Y}_{21}\bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}\right.$$

$$\left.-\frac{\operatorname{cov}(\bar{Y}_{11}\cdot\bar{Y}_{21},\bar{Y}_{12})}{\bar{\bar{Y}}_1^2\,\bar{\bar{Y}}_2}\right) + E\,\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\frac{(\bar{Y}_{12}-\bar{\bar{Y}}_1)^2}{\bar{\bar{Y}}_1^2} . \qquad (9.3)$$

This gives the bias of ${}_4\bar{\bar{Y}}_2$.

9.3. Variance of ${}_4\bar{Y}_2$

We are considering the approximate variance of ${}_4\bar{Y}_2$.

$$V({}_4\bar{Y}_2) \doteq a^2Q^2\frac{N}{N-1}\frac{1}{nPQ}[R^2\sigma_1^2+\sigma_2^2-2RQ\sigma_{12}] - 2aR\frac{N}{N-1}\frac{1}{nPQ}PQ^2\sigma_{12}$$

$$+\frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right) .$$

Proof:

As before,

$$ {}_4\bar{Y}_2 = aQ \frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}} + P\bar{Y}_{21} + bQ\bar{Y}_{22} \quad . $$

Then,

$$ V({}_4\bar{Y}_2) = a^2Q^2V\left(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}}\right) + P^2V(\bar{Y}_{21}) + (1-a)^2Q^2V(\bar{Y}_{22}) $$

$$ + 2aQP \operatorname{cov}\left(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{21}\right) + 2a(1-a)\, Q^2 \operatorname{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{22}\right) $$

$$ + 2P(1-a)Q \operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{22}) \; . $$

Let us consider $V\left(\frac{\bar{Y}_{11}, \bar{Y}_{21}}{\bar{Y}_{12}}\right)$, $\operatorname{cov}\left(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{21}\right)$, $\operatorname{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{22}\right)$.

i) $$ V\left(\frac{\bar{Y}_{11} \cdot \bar{Y}_{21}}{\bar{Y}_{12}}\right) \doteq \bar{\bar{Y}}_2^2 \left[\frac{V(\bar{\bar{Y}}_{11})}{\bar{\bar{Y}}_1^2} + \frac{V(\bar{Y}_{21})}{\bar{\bar{Y}}_2^2} + 2\frac{\operatorname{cov}(\bar{Y}_{11}\bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} + \frac{V(\bar{Y}_{12})}{\bar{\bar{Y}}_1^2} \right. $$

$$ \left. - 2\frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{12})}{\bar{\bar{Y}}_1^2} - 2\frac{\operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{12})}{\bar{\bar{Y}}_1\, \bar{\bar{Y}}_2} \right] $$

$$ \doteq R^2 \frac{N}{N-1} \frac{\sigma_1^2}{nPQ} - 2R \frac{N}{N-1} \frac{\sigma_{12}}{Pn} + \frac{\sigma_2^2}{Pn} \left(\frac{N-Pn}{N-1}\right) $$

Proof: As (9.2), we can write

$$ \frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} = \bar{\bar{Y}}_2 (1+\Delta\bar{Y}_{11} + \Delta\bar{Y}_{21} + \Delta\bar{Y}_{11}\Delta\bar{Y}_{21})(1-\Delta\bar{Y}_{12}) $$

$$ \frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} \; \frac{(\bar{Y}_{12} - \bar{\bar{Y}}_1)^2}{\bar{\bar{Y}}_1^2} $$

$$= \bar{\bar{Y}}_2(1+\Delta\bar{Y}_{11}+ \Delta\bar{Y}_{21}+ \Delta\bar{Y}_{11}\Delta\bar{Y}_{21}- \Delta\bar{Y}_{12}- \Delta\bar{Y}_{11}\Delta\bar{Y}_{12}- \Delta\bar{Y}_{21}\Delta\bar{Y}_{12}$$

$$- \Delta\bar{Y}_{11}\Delta\bar{Y}_{21}\Delta\bar{Y}_{12}) + \frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}} \frac{(\bar{Y}_{12}- \bar{\bar{Y}}_1)^2}{\bar{\bar{Y}}_1^2} .$$

If we neglect all moments higher than the second in the variance of $\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}$ then we have the above result.

ii) $$\text{cov}\,(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{21}) \doteq V(\bar{Y}_{21}) + R\,\text{cov}\,(\bar{Y}_{12}, \bar{Y}_{21}) - R\,\text{cov}\,(\bar{Y}_{21}, \bar{Y}_{12})$$

$$\doteq \frac{\sigma_2^2}{Pn} (\frac{N-Pn}{N-1}) - R \frac{N}{N-1} \frac{\sigma_{12}}{Pn} \quad \text{where } R = \frac{\bar{\bar{Y}}_2}{\bar{\bar{Y}}_1}$$

Proof:

$$\text{cov}\,(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{21}) = E\,(\frac{\bar{Y}_{11}\bar{Y}_{21}^2}{\bar{Y}_{12}}) - E\,(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}})\, E\,\bar{Y}_{21} .$$

Since,

$$\frac{\bar{Y}_{11}\bar{Y}_{21}^2}{\bar{Y}_{12}} = \bar{\bar{Y}}_2^2\,(1+\Delta\bar{Y}_{11}+ \Delta\bar{Y}_{21}^2+ \Delta\bar{Y}_{11}\Delta\bar{Y}_{21}^2+ 2\Delta\bar{Y}_{21}$$

$$+ 2\Delta\bar{Y}_{11}\Delta\bar{Y}_{21})(1-\Delta\bar{Y}_{12}) + \frac{\bar{Y}_{11}\bar{Y}_{21}^2}{\bar{Y}_{12}} (\frac{\bar{Y}_{12}- \bar{\bar{Y}}_1}{\bar{\bar{Y}}_1})^2$$

and if we neglect all moments higher than the second, then

$$E\,\frac{\bar{Y}_{11}\bar{Y}_{21}^2}{\bar{Y}_{12}} \doteq \bar{\bar{Y}}_2^2\,(1 + \frac{V(\bar{Y}_{21})}{\bar{\bar{Y}}_2^2} + 2\,\frac{\text{cov}(\bar{Y}_{11},\bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} - \frac{\text{cov}(\bar{Y}_{11},\bar{Y}_{12})}{\bar{\bar{Y}}_1^2}$$

$$- 2\,\frac{\text{cov}(\bar{Y}_{21}, \bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}) .$$

$E\,\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}$ is given in (9.3) .

Hence

$$\operatorname{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{21}\right) \doteq \bar{\bar{Y}}^2\left[1 + \frac{V(\bar{Y}_{21})}{\bar{\bar{Y}}_2^2} + \frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} - \frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{12})}{\bar{\bar{Y}}_1^2}\right.$$

$$\left. - 2\,\frac{\operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} - 1 - \frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} + \frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{12})}{\bar{\bar{Y}}_1^2} + \frac{\operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}\right]$$

$$\doteq \bar{\bar{Y}}_2^2\left[\frac{V(\bar{Y}_{21})}{\bar{\bar{Y}}_2^2} + \frac{\operatorname{cov}(\bar{Y}_{11}, \bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} - \frac{\operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2}\right]$$

$$\doteq \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) - R\,\frac{N}{N-1}\,\frac{\sigma_{12}}{Pn} \quad .$$

iii) $\operatorname{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{22}\right) \doteq \operatorname{cov}(\bar{Y}_{21}, \bar{Y}_{22}) = -\frac{\sigma_2^2}{N-1}$.

Proof:

$$\operatorname{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{22}\right) = E\left(\frac{\bar{Y}_{11}\bar{Y}_{21}\bar{Y}_{22}}{\bar{Y}_{12}}\right) - E\,\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\,E\,\bar{Y}_{22} ,$$

since

$$\frac{\bar{Y}_{11}\bar{Y}_{21}\bar{Y}_{22}}{\bar{Y}_{12}} = \frac{(\bar{\bar{Y}}_1 + \bar{Y}_{11} - \bar{\bar{Y}}_1)(\bar{\bar{Y}}_2 + \bar{Y}_{21} - \bar{\bar{Y}}_2)(\bar{\bar{Y}}_2 + \bar{Y}_{22} - \bar{\bar{Y}}_2)}{(\bar{\bar{Y}}_1 + \bar{Y}_{12} - \bar{\bar{Y}}_1)}$$

$$= \frac{\bar{\bar{Y}}_2^2\,\bar{\bar{Y}}_1}{\bar{\bar{Y}}_1}(1+\Delta\bar{Y}_{11})(1+\Delta\bar{Y}_{21})(1+\Delta\bar{Y}_{22})\,\frac{1}{1+\Delta\bar{Y}_{12}}$$

$$= \bar{\bar{Y}}_2^2\,(1+\Delta\bar{Y}_{11}+\Delta\bar{Y}_{21}+\Delta\bar{Y}_{11}\Delta\bar{Y}_{21}+\Delta\bar{Y}_{22}+\Delta\bar{Y}_{11}\Delta\bar{Y}_{22}$$

$$+\Delta\bar{Y}_{21}\Delta\bar{Y}_{22}+\Delta\bar{Y}_{11}\Delta\bar{Y}_{21}\Delta\bar{Y}_{22}-\Delta\bar{Y}_{12}-\Delta\bar{Y}_{11}\Delta\bar{Y}_{12}$$

$$- \Delta\bar{Y}_{21}\Delta\bar{Y}_{12} - \Delta\bar{Y}_{11}\Delta\bar{Y}_{21}\Delta\bar{Y}_{12} - \Delta\bar{Y}_{22}\Delta\bar{Y}_{12} - \Delta\bar{Y}_{11}\Delta\bar{Y}_{22}\Delta\bar{Y}_{12}$$

$$- \Delta\bar{Y}_{21}\Delta\bar{Y}_{22}\Delta\bar{Y}_{12} - \Delta\bar{Y}_{11}\Delta\bar{Y}_{12}\Delta\bar{Y}_{21}\Delta\bar{Y}_{22}) + \frac{\bar{Y}_{11}\bar{Y}_{21}\bar{Y}_{22}}{\bar{Y}_{12}} \left(\frac{\bar{Y}_{12}-\bar{\bar{Y}}_1}{\bar{\bar{Y}}_1}\right)^2 .$$

Hence, if we neglect all moments higher than the second,

$$E\left(\frac{\bar{Y}_{11}\bar{Y}_{21}\bar{Y}_{22}}{\bar{Y}_{12}}\right) \doteq \bar{\bar{Y}}_2^2 \left(1 + \frac{\mathrm{cov}(\bar{Y}_{11}\bar{Y}_{21})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} + \frac{\mathrm{cov}(\bar{Y}_{11}\bar{Y}_{22})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} + \frac{\mathrm{cov}(\bar{Y}_{21}\bar{Y}_{22})}{\bar{\bar{Y}}_2^2}\right.$$

$$\left. - \frac{\mathrm{cov}(\bar{Y}_{11}\bar{Y}_{12})}{\bar{\bar{Y}}_1^{\,2}} - \frac{\mathrm{cov}(\bar{Y}_{21}\bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} - \frac{\mathrm{cov}(\bar{Y}_{22}\bar{Y}_{12})}{\bar{\bar{Y}}_1\bar{\bar{Y}}_2} \right) ,$$

and $E\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\right)$ is given as in (9.3).

Therefore,

$$\mathrm{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}, \bar{Y}_{22}\right) \doteq \mathrm{cov}(\bar{Y}_{21}, \bar{Y}_{22}) .$$

Finally,

$$V({}_4\bar{Y}_2) \doteq a^2Q^2 \left[V\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\right) + V(\bar{Y}_{22}) - 2\,\mathrm{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\bar{Y}_{22}\right)\right]$$

$$- 2a\left[Q^2V(\bar{Y}_{22}) - PQ\,\mathrm{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\bar{Y}_{21}\right) - Q^2\mathrm{cov}\left(\frac{\bar{Y}_{11}\bar{Y}_{21}}{\bar{Y}_{12}}\bar{Y}_{22}\right)\right.$$

$$\left. + PQ\,\mathrm{cov}(\bar{Y}_{21}\bar{Y}_{22})\right] + P^2V(\bar{Y}_{21}) + Q^2V(\bar{Y}_{22}) + 2PQ\,\mathrm{cov}(\bar{Y}_{21}\bar{Y}_{22})$$

$$= a^2Q^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\left[R^2\sigma_1^2 - 2RQ\sigma_{12} + \sigma_2^2\right] - 2aR\,\frac{N}{N-1}\,\frac{PQ^2}{nPQ}\,\sigma_{12}$$

$$+ \frac{\sigma_2^2}{n}\left(\frac{N-n}{N-1}\right) .$$

9.4. Variance of $\bar{Y}_{22} - \bar{Y}_{23}$ in $V({}_2\bar{Y}_3)$ when $Q > \frac{1}{2}$

Referring to Figure 5.2 in 5.3.1. $\bar{Y}_{22}$ can be written as a linear combination of $\bar{Y}_{20}$ and $\bar{Y}_{23}$, where $\bar{Y}_{20}$ is a sample of mean size $(Q-P)n$ based on those units which come in a sample on second occasion only. Then

$$\bar{Y}_{22} = \frac{1}{Qn}\left((Q-P)n\,\bar{Y}_{20} + Pn\,\bar{Y}_{23}\right) = \frac{Q-P}{Q}\bar{Y}_{20} + \frac{P}{Q}\bar{Y}_{23}$$

and
$$\bar{Y}_{22} - \bar{Y}_{23} = \frac{Q-P}{Q}(\bar{Y}_{20} - \bar{Y}_{23}) .$$

The variance of $\bar{Y}_{20} - \bar{Y}_{23}$ is

$$V(\bar{Y}_{20} - \bar{Y}_{23}) = V(\bar{Y}_{20}) + V(\bar{Y}_{23}) - 2\,\text{cov}\,(\bar{Y}_{20}\bar{Y}_{23})$$

$$= \frac{\sigma_2^2}{(Q-P)n}\left(\frac{N-Qn+Pn}{N-1}\right) + \frac{\sigma_2^2}{Pn}\left(\frac{N-Pn}{N-1}\right) + 2\,\frac{\sigma_2^2}{N-1}$$

$$= \frac{N}{N-1}\;\frac{Q}{(Q-P)}\;\frac{\sigma_2^2}{Pn} \quad .$$

Therefore
$$V(\bar{Y}_{22} - \bar{Y}_{23}) = \frac{N}{N-1}\,\frac{\sigma_2^2}{nPQ}\,(Q-P) .$$

9.5. Covariance $(\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{21} - \bar{Y}_{23})$ in $V({}_2\bar{Y}_3)$ when $Q > \frac{1}{2}$

Again from Figure 5.2 in 5.3.1

$$\text{cov}\,(\bar{Y}_{31} - \bar{Y}_{32}, \bar{Y}_{22} - \bar{Y}_{23}) = \text{cov}\,\bar{Y}_{31}\bar{Y}_{22} - \text{cov}\,\bar{Y}_{31}\bar{Y}_{23} .$$

Now consider cov $(\bar{Y}_{31}, \bar{Y}_{22})$.

Since
$$\bar{Y}_{22} = \frac{Q-P}{Q}\bar{Y}_{20} + \frac{P}{Q}\bar{Y}_{23}$$

$$\text{cov}\,(\bar{Y}_{31}, \bar{Y}_{22}) = \text{cov}\,(\bar{Y}_{31}, \frac{Q-P}{Q}\bar{Y}_{20} + \frac{P}{Q}\bar{Y}_{23})$$

$$= -\frac{Q-P}{Q}\,\frac{\sigma_{23}}{N-1} + \frac{P}{Q}\,\frac{\sigma_{23}}{Pn}\left(\frac{N-Pn}{N-1}\right) = \frac{\sigma_{23}}{Qn}\,\frac{N-Qn}{N-1} \quad .$$

Hence

$$\text{cov}\ (\overline{Y}_{31} - \overline{Y}_{32}, \overline{Y}_{22} - \overline{Y}_{23}) = \frac{\sigma_{23}}{Qn}\,\frac{N-Qn}{N-1} - \frac{\sigma_{23}}{Pn}\,(\frac{N-Pn}{N-1})$$

$$= \frac{N}{N-1}\,\frac{\sigma_{23}}{nPQ}\ (P-Q)\ .$$

9.6. Variance of ${}_2\overline{Y}_{3w}$ when $Q \geq \frac{1}{2}$

$$V({}_2\overline{Y}_{3w}) = \frac{\sigma_3^2}{Qn}\ (\frac{N-Qn}{N-1}) - \frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}\,Pa_{3w}$$

Proof:

In 5.3.1,

$$V({}_2\overline{Y}_3) = \frac{\sigma_3^2}{Qn}\,(\frac{N-Qn}{N-1}) + a_3^2\,\frac{N}{N-1}\,\frac{1}{nPQ}\Big[\sigma_3^2 + \sigma_2^2(Q-P) - 2(Q-P)\sigma_{23}$$

$$+ a_2^2\sigma_2^2 + a_2^2Q^2\sigma_1^2 - 2a_2^2Q^2\sigma_{12} - 2a_2P\sigma_{23}\Big] - 2a_3\,\frac{N}{N-1}\,\frac{P\sigma_3^2}{nPQ}\ .$$

From (5)

$$a_{2w}^2(\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}) = a_{2w}^2\,\frac{P\sigma_{23}}{a_3} = a_{2w}P\sigma_{23}$$

$$\sigma_3^2 + \sigma_2^2\ (Q-P) - a_{2w}P\sigma_{23} - 2\sigma_{23}(Q-P) = \frac{P\sigma_3^2}{a_{3w}}\ .$$

Hence

$$V({}_2\overline{Y}_{3w}) = \frac{\sigma_3^2}{Qn}\,(\frac{N-Qn}{N-1}) - \frac{N}{N-1}\,\frac{\sigma_3^2}{nPQ}\,Pa_{3w}\ .$$

9.7. Variance of ${}_2\overline{Y}_\alpha$ and ${}_2\overline{Y}_{\alpha w}$ when $Q \geq \frac{1}{2}$

From ${}_2\overline{Y}_\alpha$ in 5.3.1,

$$V({}_2\overline{Y}_\alpha) = \frac{\sigma_\alpha^2}{Qn}\,(\frac{N-Qn}{N-1}) + a_\alpha^2\,\frac{N}{N-1}\,\frac{\sigma_\alpha^2}{nPQ} + \sum_{i=1}^{\alpha-2} a_\alpha^2 a_{\alpha-1}^2 \cdots a_{\alpha-i+1}^2\,\frac{N}{N-1}\,\frac{\sigma_{\alpha-i}^2}{nPQ}\,(Q-P)$$

$$+ \sum_{i=1}^{\alpha-2} a_\alpha^2 \cdots a_{\alpha-i}^2 \frac{N}{N-1} \frac{\sigma_{\alpha-i}^2}{nPQ} + a_\alpha^2 \cdots a_2^2 Q^2 \frac{N}{N-1} \frac{\sigma_1^2}{nPQ}$$

$$- 2 a_\alpha \frac{N}{N-1} \frac{P\sigma_\alpha^2}{nPQ} - 2a_\alpha^2 \frac{N}{N-1} \frac{\sigma_{\alpha(\alpha-1)}}{nPQ} (Q-P) - 2a_\alpha^2 a_{\alpha-1} \frac{P\sigma_{\alpha(\alpha-1)}}{nPQ} \frac{N}{N-1}$$

$$- 2 a_\alpha^2 a_{\alpha-1}^2 \frac{N}{N-1} \frac{\sigma_{(\alpha-1)(\alpha-2)}}{nPQ}(Q-P) - 2a_\alpha^2 a_{\alpha-1}^2 a_{\alpha-2} \frac{N}{N-1} \frac{P\sigma_{(\alpha-1)\alpha-2}}{nPQ}$$

$$\vdots$$

$$- 2a_\alpha^2 a_{\alpha-1}^2 \cdots a_{\alpha-i+1}^2 \frac{N}{N-1} \frac{\sigma_{(\alpha-i+1)(\alpha-i)}}{nPQ} (Q-P) - 2a_\alpha^2 \cdots a_{\alpha-i+1}^2 a_{\alpha-i} \frac{N}{N-1}$$

$$\cdot \frac{P\sigma_{(\alpha-i+1)(\alpha-i)}}{nPQ}$$

$$\vdots$$

$$-2a_\alpha^2 \cdots a_3^2 \frac{N}{N-1} \frac{\sigma_{23}}{nPQ} (Q-P) - 2a_\alpha^2 \cdots a_2^2 a_2 \frac{N}{N-1} \frac{P\sigma_{23}}{nPQ}$$

$$- 2a_\alpha^2 \cdots a_2^2 Q^2 \frac{N}{N-1} \frac{\sigma_{12}}{nPQ}$$

$$= \frac{\sigma_\alpha^2}{Qn} \left(\frac{N-Qn}{N-1}\right) + \frac{N}{N-1} \frac{1}{nPQ} a_\alpha^2 \Big\{ \sigma_\alpha^2 + \sigma_{\alpha-1}^2 (Q-P) - 2 \sigma_{\alpha(\alpha-1)} (Q-P) - 2a_{\alpha-1} P\sigma_{\alpha(\alpha-1)}$$

$$+ a_{\alpha-1}^2 \Big[\sigma_{\alpha-1}^2 + \sigma_{\alpha-2}^2 (Q-P) - 2\sigma_{(\alpha-1)(\alpha-2)} (Q-P) - 2a_{\alpha-2} P\sigma_{(\alpha-1)(\alpha-2)}$$

$$+ a_{\alpha-2}^2 \Big[\sigma_{\alpha-2}^2 + \sigma_{\alpha-3}^2 (Q-P) - 2 \sigma_{(\alpha-2)(\alpha-3)} (Q-P) - 2a_{\alpha-3} P\sigma_{(\alpha-2)(\alpha-3)}$$

$$\vdots$$

$$+ a_3^2 \Big[\sigma_3^2 + \sigma_2^2 (Q-P) - 2\sigma_{23} (Q-P) - 2 a_2 P\sigma_{23}$$

$$+ a_2^2 \Big[\sigma_2^2 + Q^2 (Q-P) - 2Q^2 \sigma_{12} \Big]\Big] \cdots \Big]\Big]\Big\} - 2 \frac{N}{N-1} \frac{\sigma_\alpha^2}{nPQ} P a_\alpha \quad .$$

The partial derivative of $V({}_2\bar{Y}_\alpha)$ with respect to a_i provides the optimum value of a_i in (5.21).

$$\frac{\partial}{\partial a_2} V({}_2\bar{Y}_\alpha) \Rightarrow a_{2w} = \frac{P\sigma_{23}}{\sigma_2^2 + Q^2\sigma_1^2 - 2Q^2\sigma_{12}}$$

$$\frac{\partial}{\partial a_i} V({}_2\bar{Y}_\alpha) \Rightarrow a_i(\sigma_i^2+\sigma_{i-1}^2(Q-P)-2\sigma_{i,i-1}(Q-P)-a_{i-1}P\sigma_{i,i-1})-P\sigma_{i,i+1} = 0$$

$$a_{iw} = \frac{P\sigma_{i(i+1)}}{\sigma_i^2+\sigma_{i-1}^2(Q-P)-2\sigma_{i(i-1)}(Q-P)-a_{(i-1)w}P\sigma_{i(i-1)}}$$

$$\frac{\partial}{\partial a_\alpha} V({}_2\bar{Y}_\alpha) \Rightarrow a_{\alpha w} = \frac{P\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_{\alpha-1}(Q-P)-2\sigma_{\alpha(\alpha-1)}(Q-P)-a_{(\alpha-1)w}P\sigma_{\alpha(\alpha-1)}} .$$

Substituting a_{iw} in $V({}_2\bar{Y}_\alpha)$, we have

$$V({}_2\bar{Y}_{\alpha w}) = \frac{\sigma_\alpha^2}{Qn}\left(\frac{N-Qn}{N-1}\right) - \frac{N}{N-1}\frac{\sigma_\alpha^2}{nPQ} Pa_{\alpha w} .$$

9.8. Variance of ${}_2\bar{Y}_3$ when $Q \le \frac{1}{2}$

Referring to 5.3.2, the variance of ${}_2\bar{Y}_3$ is as follows.

$$\begin{aligned}
V({}_2\bar{Y}_3) = {} & V(\bar{Y}_{32}) + a_3^2\, V(\bar{Y}_{31} - \bar{Y}_{32}) + a_3^2a_2^2V(\bar{Y}_{21} - \bar{Y}_{22}) + a_2^2a_2^2Q^2V(\bar{Y}_{11}-\bar{Y}_{12}) \\
& + 2a_3 \operatorname{cov}(\bar{Y}_{32}, \bar{Y}_{31}-\bar{Y}_{32}) + 2a_3a_2\operatorname{cov}(\bar{Y}_{32},\bar{Y}_{21}-\bar{Y}_{22}) \\
& + 2a_3a_2Q \operatorname{cov}(\bar{Y}_{32},\bar{Y}_{11} - \bar{Y}_{12}) + 2a_2^2a_2\operatorname{cov}(\bar{Y}_{31}-\bar{Y}_{32},\bar{Y}_{21}-\bar{Y}_{22}) \\
& + 2a_3^2a_2Q \operatorname{cov}(\bar{Y}_{31} - \bar{Y}_{32},\bar{Y}_{11}-\bar{Y}_{12}) + 2a_3^2a_2^2Q \operatorname{cov}(\bar{Y}_{21}-\bar{Y}_{22},\bar{Y}_{11}-\bar{Y}_{12}) \\
& + a_3^2\, V(\bar{Y}_{22}-\bar{Y}_{23}) + 2a_3\operatorname{cov}(\bar{Y}_{32},\bar{Y}_{22}-\bar{Y}_{23}) + 2a_3^2\operatorname{cov}(\bar{Y}_{22}-\bar{Y}_{23},\bar{Y}_{31}-\bar{Y}_{32}) \\
& + 2a_3^2a_2\operatorname{cov}(\bar{Y}_{22}-\bar{Y}_{23}, \bar{Y}_{21}-\bar{Y}_{22}) + 2a_3^2a_2Q \operatorname{cov}(\bar{Y}_{22}-\bar{Y}_{23},\bar{Y}_{11}-\bar{Y}_{12})
\end{aligned}$$

1. $\text{cov}\,(\bar{Y}_{31}-\bar{Y}_{32},\bar{Y}_{21}-\bar{Y}_{22}) = -\frac{N}{N-1}\frac{\sigma_{23}}{Pn}\;\frac{Q}{P}$

$$\text{cov}\,(\bar{Y}_{31}-\bar{Y}_{32},\bar{Y}_{21}-\bar{Y}_{22}) = \text{cov}\,\bar{Y}_{31}\bar{Y}_{21} - \text{cov}\,\bar{Y}_{31}\bar{Y}_{22}\;.$$

Thus, $\bar{Y}_{31}$ can be written as a linear combination of $\bar{Y}_{311}$ and $\bar{Y}_{312}$ where $\bar{Y}_{311}$ is the matched portion on the first, second and third occasions. Also $\bar{Y}_{312}$ is the matched portion on the second and third occasion only. Similarly for $\bar{Y}_{21}$, $\bar{Y}_{211}$ is the matched portion on the first and second occasions only and $\bar{Y}_{212}$ is the matched portion on all three successive occasions.

Now $\text{cov}\,(\bar{Y}_{31}\bar{Y}_{21}) = \text{cov}\,(\frac{P-Q}{P}\bar{Y}_{311} + \frac{Q}{P}\bar{Y}_{312},\;\frac{Q}{P}\bar{Y}_{211} + \frac{(P-Q)}{P}\bar{Y}_{212})$

$$= -\frac{(P-Q)}{P^2}Q\frac{\sigma_{23}}{N-1} + \frac{(P-Q)^2}{P^2}\;\frac{\sigma_{23}}{(P-Q)n}\;(\frac{N-(P-Q)n}{N-1})$$

$$-\frac{Q^2}{P^2}\;\frac{\sigma_{23}}{N-1} - \frac{(P-Q)Q}{P^2}\;\frac{\sigma_{23}}{N-1}$$

$$= -\frac{\sigma_{23}}{N-1} + \frac{N}{N-1}\;\frac{P-Q}{P^2}\;\frac{\sigma_{23}}{n}$$

$$\text{cov}\,(\bar{Y}_{31}\bar{Y}_{22}) = \text{cov}\,(\frac{P-Q}{P}\bar{Y}_{311} + \frac{Q}{P}\bar{Y}_{312},\;\bar{Y}_{22})$$

$$= -\frac{P-Q}{P}\;\frac{\sigma_{23}}{N-1} + \frac{Q}{P}\;\frac{\sigma_{23}}{Qn}\;(\frac{N-Qn}{N-1})$$

$$= \frac{\sigma_{23}}{Pn}\;(\frac{N-Pn}{N-1})\;.$$

Therefore,

$$\text{cov}\,(\bar{Y}_{31}-\bar{Y}_{32},\;\bar{Y}_{21}-\bar{Y}_{22}) - \frac{N}{N-1}\frac{\sigma_{23}}{Pn}\;\frac{Q}{P}\;.$$

2. $V(\bar{Y}_{22}-\bar{Y}_{23}) = \frac{N}{N-1}\;\frac{(P-Q)}{nPQ}\;\sigma_2^2$

This is obtained from 9.4 by changing (Q-P) to (P-Q). Next, consider $\operatorname{cov}(\bar{Y}_{22}-\bar{Y}_{23}, \bar{Y}_{21}-\bar{Y}_{22})$.

3. $\operatorname{Cov}(\bar{Y}_{22}-\bar{Y}_{23},\bar{Y}_{21}-\bar{Y}_{22}) = \operatorname{cov}\bar{Y}_{22}\bar{Y}_{21} - \operatorname{cov}\bar{Y}_{23}\bar{Y}_{21} - \operatorname{cov}\bar{Y}_{22}\bar{Y}_{22} + \operatorname{cov}\bar{Y}_{23}\bar{Y}_{22}$

where

$$\operatorname{Cov}(\bar{Y}_{23}\bar{Y}_{21}) = \operatorname{cov}\left(\frac{P-Q}{P}\bar{Y}_{212} + \frac{Q}{P}\bar{Y}_{22}, \frac{Q}{P}\bar{Y}_{211} + \frac{P-Q}{P}\bar{Y}_{212}\right)$$

$$= -\frac{(P-Q)}{P^2} Q \frac{\sigma_2^2}{N-1} - \left(\frac{Q}{P}\right)^2 \frac{\sigma_2^2}{N-1} + \frac{(P-Q)}{P^2}\frac{\sigma_2^2}{n}\left(\frac{N-(P-Q)n}{N-1}\right)$$

$$-\frac{Q(P-Q)}{P^2}\frac{\sigma_2^2}{N-1} = \frac{N}{N-1}\frac{\sigma_2^2}{Pn}\left(\frac{P-Q}{P}\right) - \frac{\sigma_2^2}{N-1}$$

$$\operatorname{Cov}(\bar{Y}_{22}\bar{Y}_{23}) = -\frac{(P-Q)}{P}\frac{\sigma_2^2}{N-1} + \frac{Q}{P}\frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right)$$

$$\therefore \quad \operatorname{Cov}(\bar{Y}_{22}-\bar{Y}_{23},\bar{Y}_{21}-\bar{Y}_{22}) = -\frac{\sigma_2^2}{N-1} - \frac{(P-Q)}{P}\frac{\sigma_2^2}{N-1} + \frac{Q}{P}\frac{\sigma_2^2}{Qn}\left(\frac{N-Qn}{N-1}\right)$$

$$-\frac{N}{N-1}\frac{\sigma_2^2}{Pn}\frac{P-Q}{P} + \frac{\sigma_2^2}{N-1} = -\frac{N}{N-1}\frac{\sigma_2^2}{nPQ}\frac{P-Q}{P}.$$

The following relations can be obtained easily.

$$\operatorname{Cov}(\bar{Y}_{31}-\bar{Y}_{32}, \bar{Y}_{11}-\bar{Y}_{12}) = -\frac{N}{N-1}\frac{P-Q}{P^2}\frac{\sigma_{13}}{n}$$

$$\operatorname{Cov}(\bar{Y}_{21}-\bar{Y}_{22}, \bar{Y}_{11}-\bar{Y}_{12}) = -\frac{N}{N-1}\frac{\sigma_{12}}{Pn}$$

$$\operatorname{Cov}(\bar{Y}_{31}-\bar{Y}_{32}, \bar{Y}_{22}-Y_{23}) = 0$$

$$\operatorname{Cov}(\bar{Y}_{22}-\bar{Y}_{23}, \bar{Y}_{11}-\bar{Y}_{12}) = \frac{N}{N-1}\frac{\sigma_{12}}{Pn}\left(\frac{P-Q}{P}\right).$$

Substituting those values in $V({}_2\bar{Y}_3)$ above, we get the variance of ${}_2\bar{Y}_3$ given in 5.3.2.