REDESIGN OF THE NIAGARA TENDER FRUIT OBJECTIVE YIELD SURVEY

J. Kovar[1]

The peach, sour cherry and the grape objective yield surveys have been carried out annually in the Niagara Peninsula since 1964 in order to forecast the magnitude of change in marketable fruit production from the previous year. Timeliness of the estimates is essential in order to enable the Ontario Tender Fruit Growers Marketing Board (OTFGMB) and the Ontario Grape Growers Marketing Board (OGGMB) to establish the marketing strategies well ahead of the harvest. This paper summarizes the major changes due to the second redesign initiated in 1982. In particular, the sample design, data collection operation and modifications of the estimation procedures are elaborated upon.

1. INTRODUCTION

The decision to switch from a list frame to an area frame survey was made in the first redesign in 1974 primarily due to the lack of an adequate list of commercial growers in the Niagara Peninsula. However, in 1981 the Ontario Ministry of Agriculture and Food (OMAF) has conducted a Tree Fruit Census and a Grape Vine Census. The availability of the census data makes it possible to redesign the survey for the second time in order to reflect the changes that the industry has undergone in the last eight years. Based on discussions with OMAF, it was decided that the census lists of growers are complete and accurate and that they contain sufficient information to form the sampling frame for the Tender Fruit Surveys. As a result, the peach, sour cherry and grape surveys will be conducted employing three independent samples selected from the 1981 OMAF census lists.

The object of all three surveys is to forecast the total amount of fruit actually sold (as fresh fruit or to processors). These forecast are made by

[1] J. Kovar, Business Survey Methods Division, Statistics Canada. This work was done while the author was in the Institutional and Agriculture Survey Methods Division, Statistics Canada.

estimating a ratio of the number of pieces of marketable fruit in the current year to the corresponding total for the previous year and applying this ratio to the previous year's figure of actual amount sold reported as a tonnage by the Ontario Fruit and Vegetable Statistics Committee. Thus an assumption of high correlation of fruit weight and fruit count must be made. Secondly due to the time lag between the surveys and the harvest, it has to be assumed that any loss of fruit between these two times is consistent from year to year.

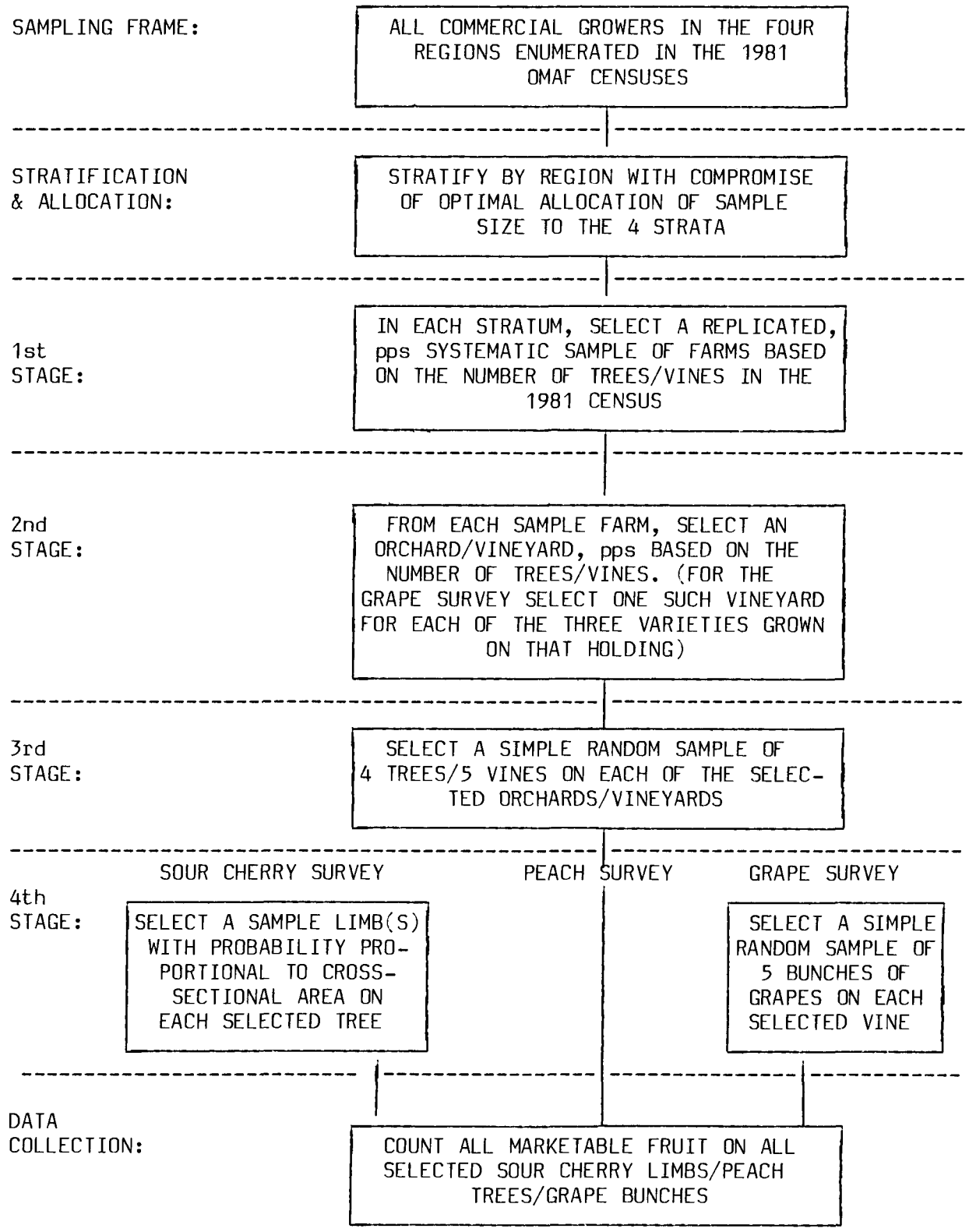
2. OVERVIEW OF THE SAMPLE DESIGN

The samples of the three objective yield surveys were selected independently according to a multistage, stratified (by geographical region), replicated, pps (farms and orchards/vineyards were selected with probability proportional to size), nearly self-weighting (all trees/vines have an approximately equal probability of selection) sample design. Figure 1 provides a visual summary of the sampling strategy. Note that due to the fact that the weight variables are collected at various points in time, the design is not exactly self-weighting.

2.1 Target Population, Sampling Frames and Total Sample Size

The target population for the three objective yield surveys comprises all commercial growers of the respective fruit in the Niagara Peninsula. Commercial growers for the three surveys were defined by OMAF as operators of those holdings which reported more than 200 peach trees, 200 sour cherry trees or 5000 grape vines respectively in the 1981 Tree Fruit or Grape Vine Censuses. Using the above definition, a separate frame was created for each of the three surveys. The lists for the peach, the sour cherry and the grape surveys contain 423, 145 and 552 commercial growers respectively. The total sample size (number of orchards/vineyards to be enumerated) for each survey was determined by OMAF's budget constraints to be in the neighbourhood of 60 for the peach survey, 55 for the sour cherry survey and 155 for the grape survey. Since for the grape survey all available varieties of interest are to be sampled on a selected farm, the final sample size for the grape survey is unknown. However, based on the 1981 Grape Vine Census, it is estimates that 62 farms will generate a sample of approximately 155 vineyards.

FIGURE 1: Sample Design for the Tender Fruit Objective Yield Surveys



2.2 Stratification and Sample Size Allocation to Regions

The Niagara Peninsula was divided into four regions for which separate estimates are required. These were defined as follows (based on the 1976 Census boundaries):

Region 1: Town of Grimsby in the Niagara Regional Municipality and township of Saltfleet in the Regional Municipality of Hamilton-Wentworth. (Township 8 of county 29 and township 4 of county 17).

Region 2: City of St. Catherines and the Town of Lincoln in the Niagara Regional Municipality. (Townships 5 and 9 of county 29).

Region 3: Town of Pelham and the Town of Thorold in the Niagara Regional Municipality. (Townships 11 and 12 of county 29).

Region 4: City of Niagara Falls and the Town of Niagara-on-the-Lake in the Niagara Regional Municipality. (Townships 3 and 10 of county 29).

Due to the increasing demand of crop production estimates by geographic area, an independent sample of farms was drawn in each of the four regions. An attempt was made to allocate the resources (i.e. number of farms sampled) optimally between regions. However, due to the unusually small population size in some regions (see Table 1) a compromise between proportional allocation, optimal allocation and a rule of "minimum of 2 farms per region per replicate" was made. The latter rule was deemed appropriate in order to diminish the possibility of complete nonresponse in a given replicate (as could be the case if only one farm per replicate was selected). The number of trees/vines in each farm was used as a measure-of-size variable for the purposes of allocation as well as for pps selection in the first and second stages. Previous results [6] indicate that other proxy variables (such as area under cultivation) are likely to be no more efficient than the tree-count variable.

TABLE 1: Population and Sample Sizes for the Tender Fruit Surveys of Commercial Growers by Region

	REGION 1		REGION 2		REGION 3		REGION 4		TOTAL	
	POPL'N	SAMPLE	POPL'N	SAMPLE	POPL'N	SAMPLE	POPL'N	SAMPLE	POPL'N	SAMPLE
PEACH	15	4	198	22	15	4	195	30	423	60
SOUR CHERRY	20	4	55	20	30	20	40	12	145	56
GRAPE	67	4	275	32	46	4	164	22	552	62

2.3 First Stage Design

Within each region, for each survey, two independent replicates of farms were selected systematically (in order to obtainm a representative sample) with probability proportional to the total number of trees/vines on the holding as of the 1981 Censuses. The total sample sizes for the two replicates are displayed in Table 1 by region. Since the two replicates are selected independently and since large farms are more likely to be selected in the sample, it is to be expected that a certain amount of overlap between replicates will exist. In fact, some farms are so large, that not only are they guaranteed to be in the sample, but they can appear more than once in the same replicate [4]). Each such appearance is treated as a separate event and one orchard/vineyard is selected without replacement every time the farm is selected. The actual number of distinct farms in the sample is therefore decreased as indicated in Table 2.

TABLE 2: Total Number of Distinct Farms in the Sample by Region

	REGION 1	REGION 2	REGION 3	REGION 4	TOTAL
PEACH	3	22	3	27	55
SOUR CHERRY	4	17	15	10	46
GRAPE	4	30	4	20	58

2.4 Second Stage Design

From the second stage on, the sampling strategies involve some field operation. Once an initial contact with the farmer is made (in the spring of 1983) it is imperative that every effort be made to obtain the respondent's co-operation. It is as this time that the farmer will be requested to aid the enumerator in listing all orchards and establishing the current size (i.e. number of trees) of each for the peach and sour cherry surveys. For the grape survey a similar listing must be prepared for each of the three varieties of interest: Concord, DeChaunac and "Other". (Note that since some varieties are grown together, one vineyard can appear on several of the lists. However, its size for a particular variety listing would be measured by the number of vines of that variety only).

On each holding, for the peach and sour cherry surveys, one orchard will be selected with probability proportional to size. For the grape survey one vineyard will be selected independently for each of the three varieties actually grown on that holding, again with probability proportional to size. It cannot be overemphasized that these procedures must be followed faithfully in order not to jeopardize the validity of the estimates. Selection procedures should be monitored carefully to ensure there is no bias in the selection towards small orchards or single variety vineyards, which admittedly would be easier to enumerate.

To avoid an overlap of orchards/vineyards on farms that are selected in both replicates or more than once in the same replicate, all orchards/vineyards for a given holding are to be selected at the same time, using a pps systematic sampling method. The assignment to replicates is to be performed at random after this selection. (Note that for the grape survey, on farms which appear in both replicates, two vineyards of each variety grown are to be selected).

2.5 Third Stage Design

Once an orchard/vineyard is selected, its current count of producing trees/vines is determined and a simple random sample of four producing trees/five producing vines is selected without replacement. The trees/vines

are marked for future identification since the same units are enumerated from year to year. (In subsequent years, if a sampled tree/vine has been destroyed, pulled up or has died, a replacement tree/vine is selected and enumerated. However, it does not contribute to the estimate until its second year in the sample). Also, each year the producing tree/vine count of selected orchards/vineyards is reestablished in order that the industry's growth (decline) can be monitored. (Note that in the grape survey this implies that only vines of the particular variety sampled are to be counted in each vineyard).

2.6 Fourth Stage Design

This stage exists only for the sour cherry and grape surveys.

2.6.1 The Sour Cherry Survey

It is operationally impossible to count all sour cherries on a selected tree. To estimate the total marketable fruit count, a sample limb (or limbs) is selected with probability proportional to the cross-sectional area of the limb. A method of selecting a limb in this way is described by Jessen [3]. It consists of selecting a limb at the initial (or primary) branching point of the trunk with probability proportional to the cross-sectional area and following the selected limb to the next branching point. This is repeated until the cross-sectional area of a subsequently selected limb is within five to fifteen percent of the primary limbs cumulative cross-sectional area total. As it is not always possible to select one such limb, in some instances two limbs will have to be enumerated. The selected limb(s) on each sample tree are then marked for future identification since the same limbs are enumerated from year to year.

2.6.2 The Grape Survey

As for sour cherries, it is equally impossible to count all marketable berries on a sample vine. Thus to estimate this total, the number of bunches of grapes (i.e. those clusters containing more than five berries) is counted and 5 bunches are selected at random without replacement in order to be enumerated.

As with the other surveys, the vines are marked and are to be visited the following year.

3. DATA COLLECTION

The actual enumeration will be performed roughly four weeks before harvest each year. It is of great importance that the selected sample vines, trees and limbs as well as the orchards and vineyards be well identified in order to enable the enumerators to complete their job in the short time available. The enumerators will be required to count all marketable fruit (i.e. excluding culls which are immature or damaged fruit that will not be harvested) on the sample peach trees and the selected cherry limbs. The fruit on the entire peach tree is counted primarily due to the fact that the fruit tends to be distributed much more unevenly on a peach tree than on a sour cherry tree [2], precluding the possibility of merely enumerating sample limbs.

For the grape survey, all berries on the five selected bunches are to be counted, excluding culls. Since most bunches are very tightly packed, this will, in most cases, involve picking the fruit. For this reason and due to time constraints, it is impossible to enumerate the entire sample vine.

4. REPLACEMENTS

Even though every attempt will be made to return to the same trees, limbs or vines in the following years, there arise cases when this is impossible. (For example, branches are sawn off, trees or vines are pulled up or are otherwise destroyed). If a sour cherry tree limb was sawn off, an attempt will be made to select another limb on the same tree using the same procedures as before. In the event that this is not possible, then just as in the case of peach trees and grape vines, a new tree/vine will be selected at random in the same orchard/vineyard. In the case that the whole orchard/vineyard has been destroyed a new orchard/vineyard will be selected on the same holding using the same procedures as described in Section 2.4. In all these cases, the newly

selected sample limbs, trees, vines, orchards or vineyards will be enumerated, however they will will not contribute to the estimate until the following year's data is collected, as only matched observations are considered.

For those hopefully rare, cases where the farmer has ceased to grow the fruit of interest entirely or where the initial contact resulted in a refusal, a third "replicate" of much smaller size was selected without replacement for each of the surveys. The procedures for selecting the orchard/vineyard and the sample of trees, limbs and vines for each replacement farm are the same as those described above. The limbs, trees and vines will be enumerated every year but will contribute to the estimate only when it is necessary to rotate one of them into the sample. The sizes of the replacement sample are indicated in Table 3 by region.

Table 3: Sample Sizes of the Replacement Sample for the Tender Fruit Survey by Region

	REGION 1	REGION 2	REGION 3	REGION 4	TOTAL
PEACH	1	2	1	3	7
SOUR CHERRY	1	2	2	2	7
GRAPE	1	3	1	2	7

5. ESTIMATION FORMULAE

5.1 Estimates of Fruit Count per Tree/Vine

Denote by y_τ the total number of marketable fruit on a tree (vine) τ.

Then for the peach survey, y_τ is estimated by $\hat{y}_\tau$, the total number of marketable peaches counted on a sample tree τ. For the sour cherry survey, y_τ is estimated by

$$\hat{y}_\tau = \hat{y}_{\tau\ell}/p_\ell \qquad (5.1.1)$$

where $\hat{y}_{\tau\ell}$ is the total number of marketable sour cherries counted on the sample limb(s) ℓ of the selected sample tree τ;

and p_ℓ is the probability of selecting the sample limb(s) ℓ. Finally for the grape survey, y_τ is estimated by

$$\hat{y}_\tau = \frac{N_\tau}{n_\tau} \sum_{\ell=1}^{n_\tau} \hat{y}_{\tau\ell} \qquad (5.1.2)$$

where N_τ is the total number of bunches of grapes on the sample vine τ;

n_τ is the number of bunches of grapes that were enumerated on the sample vine τ (typically $n_\tau = 5$);

and $\hat{y}_{\tau\ell}$ is the number of berries on a bunch ℓ of the sample vine τ.

5.2 Regional Estimates of Fruit Count by Replicate

Denote by $\hat{Y}_{ar}$ the estimated total number of marketable fruit in replicate r of region (area) a in the current year. For the grape survey,

$$\hat{Y}_{ar} = \sum_{v=1}^{3} \hat{Y}_{arv} \qquad (5.2.1)$$

where $\hat{y}_{arv}$ is the estimated total grape count of variety v in replicate r of region a.

For the purpose of uniformity of the following formulae, for the sour cherry and peach surveys $\hat{Y}_{ar}$ and $\hat{Y}_{arv}$ can be used interchangeably, since there is only one variety of sour cherries and peaches to be estimated. (In other words, the subscript v can be ignored for the sour cherry and peach surveys). Then $\hat{Y}_{arv}$ can be estimated by

$$\hat{Y}_{arv} = \frac{1}{n^*_{arv}} \sum_{f=1}^{n_{arv}} \sum_{b=1}^{n_{arvf}} \frac{N^{81}_{a}}{N^{81}_{arf}} \times \frac{N^{83}_{arvf}}{N^{83}_{arvfb}} \times \frac{N^{C}_{arvfb}}{n_{arvfb}} \sum_{\tau=1}^{n_{arvfb}} \hat{Y}_{arvfb\tau} \qquad (5.2.2)$$

with $\hat{Y}_{arvfb\tau}$ = the current year's estimated total number of marketable fruit on the tree/vine τ (variety v) in orchard/vineyard b, on farm f, in replicate r, of region a (i.e. $\hat{y}_{\tau}$);

n_{arvfb} = the number of trees/vines (of variety v) sampled in the current year in orchard/vineyard b, of farm f, in replicate r, of region a (typically $n_{arvfb} = 4$ for the sour cherry and peach surveys and 5 for the grape survey);

n_{arvf} = the number of orchards/vineyards (sampled for variety v in the current year) on farm f, in replicate r of region a (typically $n_{arvf} = 1$ except for duplicates, i.e. large farms selected more than once in the same replicate);

n_{arv} = the total number of distinct farms on which variety v was sampled in the current year) in replicate r, of region a;

n^*_{arv} = the total number of orchards/vineyards (sampled for variety v in the current year) in replicate r of region a

(i.e. $n^{*}_{arv} = \sum_{f=1}^{n_{arv}} n_{arvf}$);

N^{C}_{arvfb} = the total current count of producing trees/vines (of variety v) in orchard/vineyard b, on farm f, in replicate r, of region a;

N^{83}_{arvfb} = the total count of trees/vines (of variety v), in orchard /vineyard b, on farm f, in replicate r of region a as of the 1982/1983 mapping operation;

N^{83}_{arvf} = the total count of trees/vines (of variety v), on farm f, in replicate r, of region a as of the 1983 listing operation;

N^{81}_{arf} = the 1981 Census count of total trees/vines (all varieties) on farm f replicate r, of region a (supplied with the sample listing)

and N^{81}_{a} = the 1981 Census count of all trees/vines in region a (as per Table 4)

Table 4: 1981 Census Counts of Trees/Vines (N^{81}_{a}) for the Tender Fruit Surveys by Region

	REGION 1	REGION 2	REGION 3	REGION 4	TOTAL
PEACH	13,094	389,157	10,271	411,697	824,219
SOUR CHERRY	9,496	50,888	55,449	32,536	148,369
GRAPE	1,142,067	5,660,008	946,509	3,975,202	11,729,786

5.3 Regional Estimates of Ratio of Change and their Precision

The ratio of change in production in region a from the previous year, denoted by R_a, is estimated by

$$\hat{R}_a = \hat{Y}_a / \hat{X}_a \qquad (5.3.1)$$

where $\hat{Y}_a$ = the estimated total marketable fruit count (of peaches, sour cherries, grapes or grapes of variety v) in region a in the current year is given by

$$\hat{Y}_a = \frac{1}{2} \sum_{r=1}^{2} \hat{Y}_{ar} \qquad (5.3.2)$$

in the case of peaches, sour cherries, and <u>total</u> grapes of all varieties, and by

$$\hat{Y}_a = \hat{Y}_{av} = \frac{1}{2} \sum_{r=1}^{2} \hat{Y}_{arv} \qquad (5.3.3)$$

in the case of grapes by variety, and where $\hat{X}_a$, $\hat{X}_{ar}$, $\hat{X}_{arv}$ are the corresponding previous year's estimates.

(Note that the subscript v can now be dropped as all estimates are treated in the same manner, be it peach, sour cherry, total grape or grapes by variety estimates.)

Define the variances of $\hat{Y}_a$ and $\hat{X}_a$ and their covariance by

$$V(\hat{Y}_a) = (\hat{Y}_{a1} - \hat{Y}_{a2})^2/4 = D_{ya}^2/4 \qquad (5.3.4)$$

$$V(\hat{X}_a) = (\hat{X}_{a1} - \hat{X}_{a2})^2/4 = D_{xa}^2/4 \qquad (5.3.5)$$

$$\mathrm{Cov}(\hat{Y}_a, \hat{X}_a) = (\hat{Y}_{a1} - \hat{Y}_{a2})(\hat{X}_{a1} - \hat{X}_{a2})/4 = D_{ya} D_{xa}/4$$

where the numeric subscripts refer to the replicate number. Then the variance, $V(\hat{R}_a)$, of the ratio of change estimate, $\hat{R}_a$, can be estimated by [1].

$$\hat{V}(\hat{R}_a) = \frac{1}{\hat{X}_a^2} \{V(\hat{Y}_a) - 2\,\hat{R}_a\,\mathrm{Cov}(\hat{Y}_a,\hat{X}_a) + \hat{R}_a^2\,V(\hat{X}_a)\}$$

$$= \left\{ \frac{D_{ya}}{S_{xa}} - \frac{S_{ya}\,D_{xa}}{S_{xa}^2} \right\}^2 \qquad (5.3.7)$$

with $\quad S_{ya} = \hat{Y}_{a1} + \hat{Y}_{a2}$

$$D_{ya} = \hat{Y}_{a1} - \hat{Y}_{a2}, \text{ etc} \qquad (5.3.8)$$

The coefficient of variation of $\hat{R}_a$ is then estimated by

$$C\hat{V}(\hat{R}_a) = \frac{\{\hat{V}(\hat{R}_a)\}^{\frac{1}{2}}}{\hat{R}_a} \times 100\% \qquad (5.3.9)$$

5.4 Regional Estimates of Total Fruit Production and their Precision

Denoting by X_a^T the previous year's actual yield (tonnage) in region a and by $\hat{Y}_a^T$ the corresponding current year estimate, then $\hat{Y}_a^T$ is given by

$$\hat{Y}_a^T = X_a^T \hat{R}_a \qquad (5.4.1)$$

with its coefficient of variation estimated by

$$\hat{CV}(\hat{Y}_a^T) = \frac{X_a^T \{\hat{V}(\hat{R}_a)\}^{\frac{1}{2}}}{X_a^T \hat{R}_a} \times 100\% = \hat{CV}(\hat{R}_a) \qquad (5.4.2)$$

5.5 Estimates of Total Fruit Production and their Precision

Denote by $\hat{Y}^T$ the estimated total fruit production over all four regions in the current year. Then $\hat{Y}^T$ is given by,

$$\hat{Y}^T = \sum_{a=1}^{4} \hat{Y}_a^T \qquad (5.5.1)$$

with a coefficient of variation estimated by

$$C\hat{V}(\hat{Y}) = \frac{\{\sum_{a=1}^{4} (X_a^T)^2 V(\hat{R}_a)\}^{\frac{1}{2}}}{\hat{Y}^T} \times 100\% \qquad (5.5.2)$$

6. SUMMARY

The first enumeration will take place in 1983, however, it will not be until the summer of 1984 that the first estimates from the redesigned survey will be produced. For this reason, it will be necessary to conduct both the old and the new surveys in 1983 so that estimates will be available for that year.

Even though the survey was designed to be self-weighting, it is only approximately so due to the time differences of the 1981 Censuses, the 1983 initial listing operation and the subsequent enumerations. The estimation formulae presented in the previous section take these time differences into account. However, due to the appealing simplicity of the self-weighting estimate, an investigation of its performance has been proposed once the data becomes available.

REFERENCES

[1] Cochran, W.G. (1963). Sampling Techniques, John Wiley and Sons, New York.

[2] Davidson, G. (1977). Redesign of the sour cherry, peach and grape objective yield surveys in the Niagara Peninsula, Survey Methodology 3, 38-61.

[3] Jessen, R.J. (1955). Determining the fruit count on a tree by randomized branch sampling, Biometrics 11, 99-109.

[4] Kish, L. (1965). Survey Sampling, John Wiley and Sons, New York.

[5] Murthy, M.N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

[6] Singh, U. and Sukhatme, B.V. (1980). Sampling for estimating production of fruit crops, Sankhya C42, 17-30.