Statistics Canada Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

11-617

no. 93-11

c. 2

Canada

WORKING PAPER NO. BSMD-93-011E/F

CAHIER DE TRAVAIL NO. BSMD-93-011E/F

METHODOLOGY BRANCH

DIRECTION DE LA MÉTHODOLOGIE

#55723 (E)
#55726 (F)

C.8

SURVEY DESIGNS FOR THE 1993 CROP AND LIVESTOCK SURVEYS: LIST FRAME

by

J. Trépanier

October 1993

STATISTICS CANADA STATISTIQUE CANADA
DEC 23 1993
LIBRARY
BIBLIOTHÈQUE

SURVEY DESIGNS FOR THE 1993 CROP AND LIVESTOCK SURVEYS: LIST FRAME

J. Trépanier

ABSTRACT

Until 1992, the National Farm Survey (NFS) was the second most important agricultural survey (next to the five-yearly Census of Agriculture). It collected agricultural data on cropland areas, livestock and farmers' finances. In 1993 the NFS was replaced by three surveys that cover crop, livestock and financial variables separately. The three surveys use a common area frame and a list frame. The financial survey list frame is somewhat different from the frame used for the other two surveys. A separate sample design has been developed for each of the three surveys. This report describes the sample designs developed for the list frame and used in the new crop and livestock surveys. These designs use both univariate and multivariate stratification methods and introduce a number of measures designed to lessen respondent burden.

I. INTRODUCTION

From 1983 to 1992, most agricultural data on cropland areas, livestock and farmers' receipts and expenses in all Canadian provinces (except Newfoundland) were collected by the National Farm Survey (NFS). The annual NFS entailed two contacts: June (cropland areas) and July (livestock and financial data). A subsample of the July sample was drawn and subsequently used as the sample for the January Farm Survey (JFS), which dealt mainly with livestock data.

To complement the data on cropland areas, a series of seven surveys, commonly known as the Panel, was conducted throughout the year. These surveys covered mainly farmers' seeding intentions, stocks of grain on the farms, and crop yields.

Following a Census of Agriculture, most farm surveys are redesigned: the survey frame is updated and some aspects of the sample design are usually modified. Following the 1991 Census, the NFS was completely redesigned. There is no doubt that the survey frame required updating. More importantly, assessment of the Agriculture Statistics Program and its probable evolution over the next few years suggested that the NFS be replaced by three separate surveys on crops, livestock and financial matters.

It was decided that a new list frame and a new area frame would be constructed, and that the two frames would be common to the new surveys on livestock and crops (the financial survey uses a slightly different list frame, but the same area frame). The NFS used up to two list frames and one area frame for some provinces (Julien and Maranda, 1990). This paper discusses the list frame aspect of the new surveys on livestock and crops. Two livestock surveys are to be conducted (in July and January). The sample for the January survey will again be a subsample of the July survey. The sampling plan for these surveys is based on a multivariate stratification. A series of six crop surveys to be introduced will be distributed throughout the year. This series, although based largely on the existing Panel, will now include the crop portion of the NFS. Stratification is univariate. A master sample is drawn and divided into subsamples to control the crop survey overlap. To further lessen respondent burden, overlap of the crop master sample and the July livestock sample has been reduced using a method due to Kish and Scott and adapted to our needs.

The following sections describe in detail the methodology used in developing the sample designs for the livestock surveys and the crop surveys. The report begins by describing the creation of the list frame common to the livestock and crop surveys, and goes on to describe the sample design developed using the list frame for each type of survey. Topics discussed include stratification, sample allocation and selection. The interested reader

will find additional technical details in Trépanier (1993).

II. CREATION OF THE LIST FRAME

The crop and livestock surveys use exactly the same list frame, which covers all Canadian provinces except Newfoundland (Newfoundland reported only 725 farms in the 1991 Census of Agriculture; separate surveys are conducted there.) Because, at the time of redesign, the Census was both the most complete and most up-to-date available source, the list frame contains all Census farms in the provinces covered, with the exception of specific exclusions. These exclusions (institutional farms and farms on Indian reserves) are not contacted because of their high cost of data collection compared to their contribution to the estimates. Following the survey, adjusted estimates are generally produced to take these exclusions into account. Because an Indian reserve is defined at the Census subdivision (CSD) level, we have identified the following CSD types:

CSD TYPE =	R:	Indian reserve
	S-E:	Indian settlement
	IGD:	Indian Government District
	VC:	Cree village
	VK:	Naskapi village

All Census farms that fall into one of these CSD categories are excluded. In fact, only farms on Indian reserves (CSD = R) were found. In all, 714 farms out of 279,318 (0.3%) were excluded because of geographic location on an Indian reserve or institutional status.

Furthermore, farms owned by multiholding corporations are identified as such on the sampling frame. Although this has no impact on the sampling frame composition, the information is required for stratification. Finally, community pastures are also identified on the frame. Like multiholding corporations, they are given special treatment at the time of stratification for both crop surveys and livestock surveys. The complete list frame contains 278,604 farms of which 469 are multiholdings' farms and 464 are community pastures.

III. CROP SURVEYS

1. Overview

Starting in 1993, crop data in Canada are collected by a series of six surveys conducted at different times of the year. This series of six surveys is the result of combining the crop portion of the NFS and the series of seven crop surveys (Panel). Of the seven previous surveys, the August yield survey was eliminated, and the March surveys on seeding intentions and stocks of grain on the farms were consolidated. Table 1 shows the new series of surveys and compares their reference dates to those of the NFS and the Panel. We see that the survey on seeding intentions takes place in early spring. The stocks of grain survey cycle begins in December, when stocks on the farms are at their highest, continues during the month of March of the following year, and ends in July, when the stocks are at their lowest. The first crop yield survey (conducted in September) provides most of the final data for the Maritime provinces, Québec, Ontario and British Columbia (except Peace River district) while the November survey provides the remainder of the final data for Québec and Ontario and all the final data for the Prairie provinces and the Peace River district in British Columbia.

British Columbia is divided into two very distinct crop regions: the Peace River district, very similar to the Prairies, and the rest of British Columbia, more like Eastern Canada. For practical reasons, these two regions are treated as two separate provinces for sample design development.

The sections which follow describe the various steps in redesigning the crop surveys. The redesign is largely based on the previous crop surveys (Panel) sample design. For additional information on the subject, see Bélanger (1990).

Table 1: New Series of Crop Surveys (Starting in 1993) Compared to the NFS and the Previous Panel (up to 1992).

Up to 1992	Starting in 1993
Survey on stocks of grain December 31	Survey on stocks of grain December 31
Survey on seeding intentions March 15	Survey on stocks of grain and seeding intentions March 31
Survey on stocks of grain March 31	
National Farm Survey - cropland areas June 1	Survey on cropland areas June 1
Survey on stocks of grain July 1	Survey on stocks of grain July 1
Survey on crop yields August 15	---
Survey on crop yields September 1	Survey on crop yields September 1
Survey on crop yields[1] November 15	Survey on crop yields[1] November 15

1 This survey does not cover the Maritime provinces and British Columbia (except the Peace River district).

2. Stratification

List frame stratification is carried out independently in each Canadian province. For the crop surveys, there are two types of strata: self-representing (take-all) strata and other (take-some) strata. The sampling frame is stratified only once and stratification is the same for each survey.

2.1. Self-Representing Strata

2.1.1. Farms Owned by Multiholding Corporations and Community Pastures

Community pastures and farms owned by multiholding corporations are given special treatment at the time of data collection. Community pastures are grouped together in a single take-all stratum in each province, and the same holds for all multiholding corporation farms. For the latter, a senior officer of the corporation is contacted once only to collect the data on all components. For the crop surveys, only corporations that grow crops are contacted.

2.1.2. Farms Specified Using the Sigma-Gap Rule

The sigma-gap rule is used to identify farms that are extremely large with respect to a given variable and dissociate themselves from the rest of the population. These farms, referred to as "specified farms", are placed in a separate take-all stratum, commonly called the specified farm stratum. This practice guards against the estimates being inordinately influenced by these farms and is thus a protection against outliers.

A specified farm stratum was constructed in each province. The variable used is total cropland area in the 1991 Census. For the Maritime provinces, Québec, Ontario and British Columbia (excluding the Peace River district), total cropland area including area under hay was used. Although hay is not a key variable in our surveys, regular rotation of crops (including hay) is common among farmers in these provinces. An area under hay may next year be seeded with another crop (e.g. barley, rye), that is a key variable in our surveys. In the Canadian Wheat Board (CWB) region (Manitoba, Saskatchewan, Alberta and the Peace River district of British Columbia), because crop rotation is not a common practice, the area under hay is subtracted from total cropland area. Hay was defined as the sum of 1991 Census figures for areas of alfalfa, other tame hay, other fodder crops and forage crop seed grains. It is important to note that for each province farms owned by multiholding corporations and community pastures are excluded from the specified farm identification process as each group has already been included in its own take-all stratum and may affect the calculations described below.

The sigma-gap rule can be described as follows:

Let X_i be the value of the specification variable for unit i.

From the $X_i > 0$, the median (MED) and standard deviation (σ) are defined.

Let $X_{(1)}, X_{(2)}, \ldots, X_{(N)}$ be the ordered $X_i > 0$.

Let k, if it exists, be the smallest number such that $X_{(k)} > \text{MED}$ and $X_{(k)} - X_{(k-1)} > \sigma$, then the farms having $X_i \geq X_{(k)}$ are said to be specified.

For the crop surveys, a total of 58 farms out of 277,671 (0.02%) were designated specified.

2.2. Other Strata

2.2.1. Stratification Levels

The first level of stratification is the province. In addition to the requirement for good provincial estimates, there is an increasing demand by provincial governments for sound sub-provincial estimates. Unfortunately, as sample size is limited by budget restrictions, it is difficult to produce reliable estimates for lower levels at all times. One solution put forward was to use sub-provincial estimation areas as a second level of stratification.

A study was conducted for the province of Ontario. The province contains 14 sub-provincial estimation areas. While setting the number of strata, analysis showed that stratification at the level of the 14 areas (i.e. one stratum cannot overlap two areas), as opposed to stratification at the provincial level, would improve the coefficients of variation (CV) at the sub-provincial level; however the increase in provincial CVs was too great to justify using the method. The study did point out a compromise solution: stratification of larger sub-provincial areas produced by consolidating sub-provincial estimation areas. With this procedure, provincial CVs are compromised less seriously and CVs for sub-provincial estimation areas are better than those obtained by single stratification at the provincial level. In Trépanier (1993), the sub-provincial estimation areas and the consolidations made to produce sub-provincial areas that we call "stratification areas" are defined on the basis of Census divisions and subdivisions. These consolidations were suggested by Agriculture Division, which attempted to group homogeneous agricultural areas together as much as possible. Five stratification areas were formed in Ontario. Similar procedures were used in other provinces when necessary.

2.2.2. Number of Strata

A minimum of four strata per stratification area were created. The size, in terms of number of farms, and total cropland area (minus hay in the West) were calculated for each area. More strata (maximum 10) were produced for areas where the number of farms or cropland area was significant.

2.2.3. Stratification Method

Farms not included in one of the self-representing strata described in section 2.1 were stratified within each province and each stratification area. A univariate stratification method was used, with the stratification variable being total cropland area minus hay in the CWB region and total cropland area elsewhere. The stratification variable is the same as the one used to identify specified farms in order to keep the univariate aspect of stratification. Stratum limits were calculated by means of the Sethi (1963) algorithm used in conjunction with Neyman's optimum sample allocation. Stratum boundaries and sizes can be found in Trépanier (1993).

3. Master Sample

The crop survey series includes six surveys in all. To lessen respondent burden as much as possible, a master sample was drawn. The master sample was divided into a number of disjoint subsamples of equal size in each province. The number and size of subsamples were determined in a manner that ensures that each survey can use a set of these subsamples whose combined sample size is near the desired sample size. If an overlap of the samples for two surveys is desired, common subsamples are selected; if not, different subsamples are selected. The master sample is drawn from the "other" strata (see Chapter III, Section 2.2). The farms in the self-representing strata are added to the samples for each survey. The following sections describe the various steps in selecting the master sample and dividing it into subsamples.

3.1. Sample Sizes for Crop Surveys at the Large Region level

For crop survey purposes, Canada is divided into three large regions:

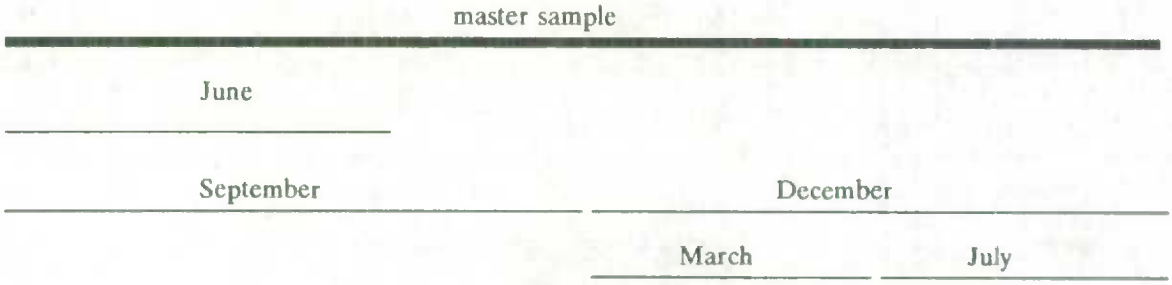
- Maritimes and British Columbia (excluding Peace River district).
- Québec and Ontario.
- CWB region.

Figure 1 shows the distribution of the master sample among the crop surveys. For instance, it shows that the samples for the March survey and the July survey form the sample for the December survey. The procedure allows for production of specific longitudinal statistics. Table 2 shows desired sample sizes by survey and large region as provided by Agriculture Division.

Figure 1: Distribution of the Master Sample among the Crop Surveys.

(not to scale)

(i) Maritimes - British Columbia (excluding Peace River district)



(ii) Québec and Ontario

master sample

June | September | December

November | March | July

(June or September farms that reported specialty crops)

(iii) CWB region

master sample

June | September | December

November | March | July

(units that reported specialty crops in June plus an additional sample)

Table 2: Desired Sample Size for Each Crop Survey by Large Region

SURVEY	REGION		
	Maritimes + B.C. (excluding Peace River)	Québec + Ontario	CWB
Intentions-Stocks (March)	400	2,500	5,000
Areas (June)	600	6,800	16,000
Stocks (July)	400	3,000	6,000
Yield (September)	1,000	7,500	8,250
Yield (November)	-	9,000[1]	16,250[2]
Stocks (December)	800	5,500	11,000

1 This figure is approximate as the November sample is in fact made up of farms that reported specialty crops in the June or September survey.

2 This figure includes a sample of 13,750 farms as well as the farms that reported specialty crops in the June survey. The size of this last group is estimated at 2,500 farms.

3.2. Subsample Size

3.2.1. Original Size

As stated above, we wished to divide the master sample into subsamples of equal size in a given large region R. In fact it was virtually impossible to find a reasonable subsample size T_R such that, for any given survey i, $x_{iR} * T_R = n_{iR}$ where n_{iR} is the desired sample size for survey i in R and x_{iR} the integer number of subsamples used of size T_R. A trivial solution to this problem is of course $T_R = 1$, but we were looking for a fairly large T_R. We were prepared to accept a variation of up to 15% between the desired size and the actual size.

We were looking for T_R such that:

$$\frac{|X_{iR} \times T_R - n_{iR}|}{n_{iR}} \times 100 \leq 15, \forall i \epsilon R.$$

This algorithm was applied for each of the large regions defined above. Although, in many cases, a number of solutions were available, we selected the one that gave the smallest number of subsamples, i.e. subsamples that were largest in size.

3.2.2. Limited Follow-up Method

At the time of the redesign, it was planned to follow-up farms during the data collection until the response rate was approximately 80% (the rate varies by survey and large region) in order to limit the cost. The subsample size was increased by 25% to ensure that the expected sample of respondents was nearly equal to the desired sample size. Because the bias caused by this approach is difficult to measure, this method was abandoned and it was decided that all farms would be contacted up to 8 times in order to solicit a response. Response rates in excess of 80% have been realized on all occasions, so far. As a result, the sample sizes may have to be reviewed prior to the next redesign.

3.2.3. Additional Subsamples

One more subsample was created in the Maritime provinces and British Columbia (excluding Peace River district); two more subsamples were created in the other two large regions. For the time being, these subsamples will not be used in the crop surveys. They may be used in the future if an additional sample is required.

3.3. Allocation and Provincial Sizes of the Master Sample

Once the subsample size per large region is defined, it must be allocated to the provinces that make up the region. This is done on the basis of provincial sample sizes in the 1991 National Farm Survey in order to keep the distribution similar to the previous cycle of surveys. In this way the relative workload by province will remain the same and as well as because the allocation had proved to be good for both crops and livestock estimates.

Let	n_R =	size of subsample in region R
	n_p =	size of subsample for province p in R
	n_{R_NFS} =	1991 NFS sample size in region R
	n_{p_NFS} =	1991 NFS sample size in province p in R

Thus $n_p = n_R * n_{p_NFS} / n_{R_NFS}$ rounded off to the nearest integer.

Table 3 shows subsample allocation and master sample sizes.

Table 3: Master Sample Size by Province.

PROVINCE	1991 NFS	SIZE OF ONE SUBSAMPLE	NUMBER OF SUBSAMPLES	SIZE OF MASTER SAMPLE
P.E.I.	487	56	6	336
Nova Scotia	571	66	6	396
N.B.	553	64	6	384
B.C. (except Peace River)	2,095	243	6	1,458
TOTAL	3,706	429	6	2,574
Québec	6,478	625	16	10,000
Ontario	11,193	1,080	16	17,280
TOTAL	17,671	1,705	16	27,280
Manitoba	5,079	783	20	15,660
Saskatchewan	9,609	1,482	20	29,640
Alberta	7,046	1,087	20	21,740
B.C. (Peace River)	371	57	20	1,140
TOTAL	22,105	3,409	20	68,180

3.4. Master Sample Allocation to the Strata

Optimal allocation that minimized the CV of the stratification variables was used. A computer program obtained from the United States Department of Agriculture (USDA) and already available was used in allocating the provincial master sample size to the strata. The program allocates a given sample size in a manner that minimizes a linear combination of the squared coefficients of variation of the selected allocation variables ($\Sigma\, a_i CV_i^2$). The CV_is referred to here are the theoretical CVs calculated from the list frame data (1991 Census data) for given allocation variables i. The variable used for allocation was the one used for identification of specified farms and for stratification. Coefficient a_1 was set to 1. Thus the function to be minimized is CV_1^2, i.e. CV_1.

3.5. Selection of Master Sample

To permit future linkage to tax data, the population units included in the 1991 tax records population (i.e. farms that filled up a tax report in 1991) were identified in the population. The master sample was selected using circular systematic sampling in each stratum: the first step was a sort by estimation areas, 1991 Census divisions and subdivisions, an indicator variable of the tax record link (equals 1 if the farm is part of the 1991 tax records population and 0 otherwise) and, finally, by a random number previously allocated to each farm. The sort ensures sample representativeness in each domain of estimation, and a proportion of master sample units linked to tax records similar to that in our population. Two samples were drawn. The better of the two, according to analysis of selected statistics (e.g. sample representativeness, performance in overlap minimization (see Chapter IV, Section 3.5)) was chosen in each province.

3.6. Partition of Master Sample into Subsamples

The master sample drawn for each province was randomly partitioned into x subsamples of size T, where x and

T are as shown in Table 3. Representativeness of strata, estimation areas, Census divisions and subdivisions and links to tax records was ensured: a range of statistics was calculated to ensure that each subsample represents the population satisfactorily. Provincial sample sizes for the surveys are given in Table 4.

Table 4: Sample Size by Survey and Province.

PROV.	SAMPLE SIZE BY SURVEY[1] (number of subsamples used)					
	MARCH	JUNE	JULY	SEPT.	NOV.	DEC.
P.E.I.	56 (1)	112 (2)	56 (1)	168 (3)	-	112 (2)
N.S.	66 (1)	132 (2)	66 (1)	198 (3)	-	132 (2)
N.B.	64 (1)	128 (2)	64 (1)	192 (3)	-	128 (2)
B.C. (except Peace River	243 (1)	486 (2)	243 (1)	729 (3)	-	486 (2)
Québec	1,250 (2)	3,125 (5)	1,250 (2)	3,125 (5)	Farms with specialty crops in June or Sept.	2,500 (4)
Ontario	2,160 (2)	5,400 (5)	2,160 (2)	5,400 (5)		4,320 (4)
Manitoba	1,566 (2)	4,698 (6)	1,566 (2)	2,349 (3)	3,915[2] (5)	3,132 (4)
Sask.	2,964 (2)	8,892 (6)	2,964 (2)	4,446 (3)	7,410[2] (5)	5,928 (4)
Alberta	2,174 (2)	6,522 (6)	2,174 (2)	3,261 (3)	5,435[2] (5)	4,348 (4)
Peace River	114 (2)	342 (6)	114 (2)	171 (3)	285[2] (5)	228 (4)

1 These sizes do not include specified farms, community pastures and multiholding corporations.
2 To this number are added the farms that reported specialty crops in June.

IV. LIVESTOCK SURVEYS

1. Overview

Until 1992, the National Farm Survey (NFS) was mandated to collect data on livestock, crops and farmers' receipts and operating expenses. The livestock data were collected in two stages. In June, when the sampled farms were first contacted regarding crops, a few questions on livestock were included. In July each farm identified as a livestock producer was re-contacted and questioned in greater detail about livestock. A subsample of the July NFS sample was contacted again in January for the January Farm Survey (JFS). At that time the farmers were asked to answer a series of questions dealing mainly with livestock.

The current redesign proposes retaining the two main surveys described above. The July Livestock Survey is conducted in June (reference date July 1). The January Livestock Survey is to replace the former JFS, with the sample again a subsample of the sample from the preceding July. The most significant changes made by the redesign may be found in the characteristics of the sample design. The survey list frame will be the same as for the crop surveys, but sample design (for example, stratification) is to be developed from livestock variables rather than a broad range of agricultural variables as was done for the NFS.

2. Stratification

2.1. Self-Representing Strata

2.1.1. Farms Owned by Multiholding Corporations and Community Pastures

Following the same procedure as for the crop surveys, all farms in the same province owned by a multiholding corporation are grouped together in one stratum. Of these farms, only multiholding corporations that are livestock producers are to be contacted. Community pastures in the same province are also grouped together in a single stratum.

2.1.2. Farms Specified by the Sigma-Gap Rule

The sigma-gap rule is again used to identify specific farms important for the following variables: beef cows, milk cows, sows, total cattle, total pigs and total sheep. More specifically, the sigma-gap rule is applied to all farms in the same province (excluding community pastures and multiholding corporations) for each of the six variables listed above. The farms identified for at least one of these variables are grouped together in a stratum known as the "specified farm" stratum. For the livestock surveys, the specified farm strata include 223 farms, of which only 8 overlap the specified farm strata of the crop surveys. The very low figure (8) appears to indicate that a farm with very high crop production does not necessarily have very high livestock production.

2.2. Zero-Livestock Stratum

According to the Census data, many farms on our frame produce no livestock (cattle, pigs, sheep). However these farms are not excluded from the target population as they may begin livestock production during the period of this sampling design or they may have been incorrectly classified in the Census. For these reasons, the farms referred to as "zero-livestock" (i.e. no cattle, pigs or sheep) are grouped together in a single stratum for each province. As shown below, the sampling rate for this stratum is low.

2.3. Other Strata

In each province, all farms not included in one of the preceding strata (self-representing, zero-livestock) are stratified using the method described in the following sections.

2.3.1. Stratification Variables

Stratification is multivariate and the variables used are the same as those used to identify specified farms (see Section 2.1.2), except that they are standardized (mean = 0, variance = 1). This avoids the dominance of one large variable (e.g., total cattle) over a smaller variable. Before this variable transformation method was adopted, a number of other methods were tested, including the use of original, untransformed variables; principal components of the six original livestock variables; and variables transformed by dividing the original variable by the mean of the contributors to that variable. None of these transformation methods proved more satisfactory than the standardization.

2.3.2. Stratification Procedures

Farms in the same province were stratified using SAS FASTCLUS and CLUSTER multivariate procedures. For more detail on these procedures, see SAS Institute Inc. (1985). Because the number of farms to be stratified was so large, we first grouped the original farms into 150 clusters using FASTCLUS (10 iterations maximum). We next used CLUSTER to combine the 150 clusters, which are in effect the units to be stratified, to form the strata using a hierarchical stratification that progressively reduces the 150 units to one stratum. The Ward (1963) minimum variance method was used in stratification, with slight modification of the stratification variables: the

values used were the mean values of the original stratification variables (standardized variables) for a given cluster. Specifically, let $i = 1,2,...,N_c$ be the units of cluster c and let $c = 1,...N$ be the clusters produced by FASTCLUS. Let X be the original stratification variable (standardized livestock variable, mean = 0, variance = 1).

Thus the new stratification variable used by CLUSTER is $\bar{X}$ where

$$\bar{X}_c = \frac{1}{N_c}\sum_{i=1}^{N_c} X_i, c = 1,...,N.$$

2.3.3. Number of Strata per Province

Selection of number of strata is based on two specific but opposing objectives: reduce coefficients of variation (by increasing the number of strata) for the important variables and, in the long term, maintain a robust stratification (by decreasing the number of strata) that reduces the number of units that move from one stratum to another over a 5-year period (lifespan of the new sampling design). It is difficult to evaluate stratification robustness until a survey has been conducted. Therefore CVs were calculated for different numbers of strata with a provincial sample size being fixed and allocated to the strata using the method described in section 3.3. This exercise allowed us to identify the minimum number of strata that would give an acceptable precision.

The SAS TREE procedure was used to extract the desired number of strata from the hierarchical stratification produced by CLUSTER. Because the difference between h and h+1 strata is small, we extracted different stratifications at intervals of 2 or 3 strata. Census data were used to calculate the theoretical coefficients of variation using the following formula:

Let X be the variable of interest and x_i be the value of that variable for unit i on the list frame (i.e the 1991 Census value). Also let

H = number of strata in a given province
N_h = number of units in stratum h
n_h = number of sampled units in stratum h

$$CV(X) = \frac{\sqrt{\sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \sum_{i=1}^{N_h} (x_i - \bar{x})^2}}{\sum_{h=1}^{H}\sum_{i=1}^{N_h} x_i}.$$

The total number of strata (including self-representing strata and zero-livestock stratum) ranges between 15 and 46 from province to province.

3.July 1992 Dummy Sample

3.1. Overview

The first livestock survey conducted under the new design was the January 1993 survey. This raised a problem since the January sample is to be a subsample derived from the July sample of the preceding year. As a solution, a dummy sample was drawn for July '92 even though no survey was conducted (July '92 was covered by the

NFS). The dummy sample was not used, except to extract a subsample for the January '93 survey. A rotation of both samples is planned for subsequent years. In July '93, the July '92 dummy sample, modified by rotation, will be used.

3.2. Sample Allocation and Provincial Size

The budget allows for a sample of approximately 29,750 units for July. The proportion of the total sample allocated to a province is based on figures from the 1991 NFS (see Chapter III, Section 3.3). Provincial sample sizes are shown in Table 5.

3.3. Sample Allocation to the Strata

The sampling rate for the zero-livestock stratum was set at 1/60. If, from the figures given in Table 5 for a given province, we subtract the total number of specified farms plus community pastures plus farms sampled in the zero-livestock stratum, the remainder represents the sample to be allocated to the other strata. Multiholding corporations are not included in these calculations as their number, as yet undetermined, is very small. Allocation of the sample to the other strata is conducted in exactly the same manner as for the crop surveys (see Chapter III, Section 3.4) on the basis of the variables used to identify specified farms. Coefficients a_1, a_2, ..., a_6 were all set to 1, since all the variables were considered to be equally important.

3.4. Selection of July 1992 Dummy Sample

The July '92 dummy sample was selected using circular systematic sampling. Before selecting the appropriate number of units in a stratum, the units were sorted by Census divisions and subdivisions, tax link indicator variable and, finally, by a number generated at random for each unit. Two samples were again drawn. The better of the two, on the basis of specific statistics analyzed (e.g. sample representativeness, performance in overlap minimization (see Chapter IV, Section 3.5)), was selected for each province.

Table 5: Provincial Sample Sizes for the July 1992 Dummy Livestock Sample Excluding Multiholding Corporations.

PROVINCE	July 1992 Dummy Sample
Prince Edward Island	335
Nova Scotia	393
New Brunswick	381
Québec	4,420
Ontario	6,801
Manitoba	3,504
Saskatchewan	5,583
Alberta	6,631
British Columbia	1,703
TOTAL	29,751

3.5. Overlap Minimization - July 1992 Dummy Sample and Crop Survey Master Sample

To distribute response burden, overlap between the July 1992 dummy sample and the crop survey master sample was minimized using the Kish and Scott (1971) method for equal initial probabilities. We adapted this method to our needs. At each intersection of a livestock survey stratum and a crop survey stratum, the units included in both samples are identified. For the livestock sample, these units are replaced at random by units available in the intersection. The units originally included in both samples remain in the crop master sample only. A more detailed description of the method follows. An interested reader could find in Perry, Burt and Iwig (1993) two methods for reducing the overlap between two or more survey samples. The first method reduces the overlap using integer linear programming while the second is the method described below but generalized to k surveys ($k \geq 2$).

Let $U_{c.}$ and $U_{.l}$ be the crop and livestock strata respectively. Also let $U_{cl} = U_{c.} \cap U_{.l}$ and

s_{cl} = units in both crop and livestock samples
n_{cl} = number of units in s_{cl}
U'_{cl} = units of U_{cl} that are in neither sample
N'_{cl} = number of units in U'_{cl}

Two courses of action are possible.

(1) If $n_{cl} \leq N'_{cl}$, n_{cl} units are selected from U'_{cl} by simple random sampling. These selected units then replace the s_{cl} units in the livestock sample.
(2) If $n_{cl} > N'_{cl}$, N'_{cl} units are selected from s_{cl} by simple random sampling. These selected units are replaced by the U'_{cl} units for the livestock sample.

During overlap minimization within a given U_{cl} intersection, one unit is not necessarily replaced by another unit with the same characteristics (e.g. Census division and subdivision), or linked to a tax record if the original unit was so linked. Two samples each were drawn for livestock and for crops. Minimization of overlap was performed on each combination of one livestock and one crop samples in each province. The worst combination gave a final overlap of 4,799 units once the method applied. Based on the performance of each combination and statistics for each sample, one sample was chosen for the crop and livestock surveys. Using the method described above, the overlap between the chosen livestock dummy sample and crop master sample was reduced by 58% for a final overlap of 4,631 units.

4. January 1993 Sample

4.1. Sample Allocation and Provincial Sizes

The sample for the January livestock survey is a subsample of the July sample. The budget allows for a January sample size of 13,500 units. Allocation of the national sample size to the provinces is based on sample allocation in the 1992 January Farm Survey (JFS). In the Maritimes, the January sample is the same as the July sample. The final sizes are given in Table 6.

4.2. Sample Allocation to the Strata

Stratification for the January Livestock Survey is the same as for the July Livestock Survey. The July sample in self-representing strata and zero-livestock strata is retained in January. This means that portions of the sample sizes given in Table 6 are absorbed by these strata. The provincial size remaining from the January sample is allocated independently of July sample size. For a given stratum, the size allocated to that stratum is obviously larger for July than for January, as the total sample is at least as large. The USDA program was again used for the allocation exercise. The allocation variables are the same as for July.

4.3. Sample Selection

The number of allocated units per stratum is selected among the units sampled in July for the same stratum. Within the stratum, units are sorted by Census divisions and subdivisions, tax link indicator variable and a number generated at random for each unit. The selection is made using circular systematic sampling.

Table 6: Provincial Sample Size for January Livestock Survey, Excluding Multiholding Corporations.

PROVINCE	PROVINCIAL SAMPLE SIZE
Prince Edward Island	335
Nova Scotia	393
New Brunswick	381
Quebec	2,297
Ontario	2,718
Manitoba	1,349
Saskatchewan	2,385
Alberta	2,656
British Columbia	988
TOTAL	13,502

V. CONCLUSION

The new area frame developed for the crop and livestock surveys is not discussed here. A working paper on the area frame redesign is expected to be released soon. We should be in a position to qualify and quantify the impact of the new redesign on the crop and livestock surveys in the coming years. The sample design is now defined specifically for each type of survey, unlike the NFS which considered three types of agricultural data concurrently. This redesign, better adapted to individual survey needs, should improve effectiveness in many ways at a cost of increased burden and loss of crop, livestock and financial links.

VI. BIBLIOGRAPHY

Bélanger, Y. (1990), "Redesign of the Crop Surveys at Statistics Canada", Statistics Canada Methodology Branch Working Paper No. BSMD-90-014E/F.

Julien, C. and Maranda, F. (1990), "Sample Design of the 1988 National Farm Survey", *Survey Methodology*, 16, 117-129.

Kish, L. and Scott, A. (1971), "Retaining Units after Changing Strata and Probabilities", *Journal of the American Statistical Association*, 66, 461-470.

Perry, C., Burt, J. and Iwig, B. (1993), "Methods of Selecting Samples in Multiple Surveys to Reduce Respondent Burden", *Proceedings of the International Conference on Establishment Surveys, American Statistical Association*, to appear.

Sas Institute Inc. (1985), *SAS User's Guide: Statistics,* Version 5, Cary, North Carolina: SAS Institute Inc.

Sethi, V. K. (1963), "A Note on Optimum Stratification of Populations for Estimating the Population Means", *Australian Journal of Statistics*, 5, 20-33.

Trépanier, J. (1993), "Remaniement des enquêtes agricoles sur les commodités: base liste", Statistics Canada Technical Report.

Ward, J. H. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236-244

Statistics Canada Statistique Canada

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canada

WORKING PAPER NO. BSMD-93-011E/F

METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. BSMD-93-011E/F

DIRECTION DE LA MÉTHODOLOGIE

LES PLANS DE SONDAGE DES ENQUÊTES SUR LES CULTURES ET SUR LE BÉTAIL DE 1993: BASE LISTE

par

J. Trépanier

Octobre 1993

10101525 46
STATISTICS CANADA LIBRARY
BIBLIOTHÈQUE STATISTIQUE CANADA
Ca 003

LES PLANS DE SONDAGE DES ENQUÊTES SUR LES CULTURES ET SUR LE BÉTAIL DE 1993: BASE LISTE

J. Trépanier

RÉSUMÉ

L'Enquête nationale sur les fermes (ENF) était, jusqu'en 1992, l'enquête agricole la plus importante après le Recensement de l'agriculture mené à tous les cinq ans. Elle couvrait à la fois les données agricoles sur les superficies en cultures, le bétail et les finances des fermiers. En 1993, l'ENF a fait place à trois enquêtes distinctes traitant respectivement des variables de cultures, de bétail et des variables financières. Les trois enquêtes utilisent une base aréolaire commune et une base liste. Seule l'enquête financière a une base liste légèrement différente des deux autres enquêtes. Un plan d'échantillonnage a été développé pour chaque enquête. Le présent rapport décrit les plans d'échantillonnage associés à la base liste pour les nouvelles enquêtes sur les cultures et sur le bétail. Ceux-ci font intervenir des méthodes de stratification univariée et multivariée selon le cas, et prévoient plusieurs façons de réduire le fardeau du répondant.

I. INTRODUCTION

De 1983 à 1992, les données agricoles sur les superficies en culture, le bétail et les revenus et dépenses des fermiers des provinces canadiennes (sauf Terre-Neuve) étaient principalement recueillies par le biais de l'Enquête nationale sur les fermes (ENF). L'ENF, menée annuellement, se faisait en deux contacts: juin (superficies en cultures) et juillet (bétail et données financières). Un sous-échantillon de l'échantillon de juillet était sélectionné et constituait l'échantillon de l'Enquête de janvier sur les fermes (EJF). L'EJF s'intéressait principalement aux données sur le bétail.

Afin de compléter les données sur les superficies en cultures, une série de sept enquêtes réparties tout au long de l'année était menée. Cette série était communément appelée le "Panel". Ces enquêtes s'intéressaient surtout aux intentions d'ensemencement des fermiers, aux stocks de grains sur les fermes et aux rendements des cultures.

Après un Recensement agricole, la majorité des enquêtes agricoles sont remaniées. La base de sondage est mise à jour et on profite généralement de l'occasion pour modifier certains aspects du plan d'échantillonnage. Ainsi, suite au Recensement de 1991, un remaniement complet de l'ENF a été entrepris. La mise à jour de la base de sondage était certainement devenue nécessaire, mais bien plus important encore, l'évaluation du Programme des statistiques agricoles et de sa probable évolution dans les prochaines années suggéraient que l'ENF soit remplacée par trois enquêtes séparées: cultures, bétail et finances.

Il fut décidé qu'une nouvelle base de sondage liste et une nouvelle base aréolaire seraient construites et que ces deux bases seraient communes aux nouvelles enquêtes sur le bétail et les cultures. (L'enquête financière a une base liste légèrement différente, mais la même base aréolaire.) L'ENF comptait pour sa part jusqu'à deux bases listes et une base aréolaire pour certaines provinces (Julien et Maranda, 1990). Le présent document ne traitera que de l'aspect base liste des nouvelles enquêtes sur le bétail et les cultures. Deux enquêtes sur le bétail auront lieu: en juillet et en janvier. Encore une fois, l'échantillon de l'enquête de janvier sera un sous-échantillon de l'enquête de juillet. Le plan d'échantillonnage de ces enquêtes repose sur une stratification multivariée. D'autre part, une série de six enquêtes sur les cultures seront mises en place et réparties tout au long de l'année. Cette série s'inspire grandement du "Panel", mais intègre également la partie cultures de l'ENF. La stratification est de type univariée. Un échantillon-maître est tiré et partitionné en sous-échantillons. Cette pratique permet de contrôler le chevauchement entre les différentes enquêtes sur les cultures. Afin de diminuer encore plus le fardeau du répondant, le chevauchement entre l'échantillon-maître de cultures et l'échantillon de juillet pour le

bétail a été réduit grâce à l'adaptation d'une méthode de Kish et Scott (1971).

Les sections qui suivent dressent un portrait complet de la méthodologie employée dans l'élaboration du plan d'échantillonnage des enquêtes sur le bétail et des enquêtes sur les cultures. Dans un premier temps, ce rapport traite de la création de la base de sondage liste commune aux enquêtes sur le bétail et sur les cultures. Par la suite, le plan d'échantillonnage associé à la base liste est décrit pour les deux types d'enquêtes. Des sujets tels la stratification, la répartition et la sélection de l'échantillon y sont discutés. Le lecteur intéressé pourra trouver des détails techniques supplémentaires dans Trépanier (1993).

II. CRÉATION DE LA BASE DE SONDAGE LISTE

Les enquêtes sur les cultures et sur le bétail utilisent exactement la même base de sondage liste. Celle-ci couvre toutes les provinces canadiennes à l'exception de Terre-Neuve. (Terre-Neuve n'avait que 725 fermes au Recensement de l'agriculture de 1991. Des enquêtes séparées y sont menées.) Le Recensement étant à la fois la source la plus complète et la plus à jour disponible au moment du remaniement, la base de sondage liste contient ainsi la totalité des fermes du Recensement des provinces couvertes à l'exception de certaines exclusions. Ces exclusions sont les fermes institutionnelles et les fermes sur les réserves indiennes. Ces fermes ne sont pas contactées en raison du coût élevé associé à la collecte de leurs données comparativement à leur contribution aux estimations. Suite à l'enquête, des estimations dites ajustées sont habituellement produites pour tenir compte de ces exclusions. Sachant qu'une réserve indienne est définie au niveau de la subdivision du Recensement (SDR), nous avons identifié les types de SDR suivants:

TYPE DE SDR =	R :	réserve indienne
	S-E :	établissement indien
	IGD :	Indian Government District
	VC :	Village Cri
	VK :	Village Naskapi

Toute ferme du Recensement tombant dans un de ces SDR a été exclue. Par contre, seules des fermes sur des réserves indiennes (SDR = R) ont été trouvées. Au total, 714 fermes sur 279 318 (0,3%) ont été exclues en raison de leur situation géographique sur une réserve indienne ou de leur statut institutionnel.

D'autre part, les fermes faisant partie de corporations à opérations multiples ont dû être identifiées sur la base de sondage. Ceci ne touche en rien la composition de la base de sondage, mais apporte une information qui sera nécessaire lors de la stratification. Finalement, les pâturages communautaires ont également été identifiés sur la base. Tout comme les corporations à opérations multiples, ils recevront un traitement spécial lors de la stratification que ce soit pour les enquêtes sur les cultures ou les enquêtes sur le bétail. La base liste comprend 278 604 fermes parmi lesquelles 469 font partie de corporations à opérations multiples et 464 sont des pâturages communautaires.

III. ENQUÊTES SUR LES CULTURES

1. Aperçu

Depuis 1993, les données sur les cultures au Canada proviennent d'une série de 6 enquêtes menées à différents moments durant l'année. Cette série de 6 enquêtes est le résultat d'une combinaison de la partie cultures de l'ENF et de la série de 7 enquêtes sur les cultures ("Panel"). De ces 7 anciennes enquêtes, l'enquête sur le rendement du mois d'août a été retirée. De plus, l'enquête sur les intentions d'ensemencement de mars et l'enquête sur les stocks de grains du même mois ont été jumelées. Le tableau suivant présente la nouvelle série d'enquêtes et leur date de référence comparativement à la série précédente (le "Panel") et à l'ENF.

Tableau 1: Nouvelle série des enquêtes sur les cultures (à partir de 1993) comparée à l'ENF et à l'ancien "Panel" (jusqu'à 1992).

Jusqu'à 1992	À partir de 1993
enquête sur les stocks de grains 31 décembre	enquête sur les stocks de grains 31 décembre
enquête sur les intentions d'ensemencement 15 mars	enquête sur les stocks de grains et les intentions d'ensemencement 31 mars
enquête sur les stocks de grains 31 mars	
Enquête nationale sur les fermes - Superficies en culture 1er juin	enquête sur les superficies en culture 1er juin
enquête sur les stocks de grains 1er juillet	enquête sur les stocks de grains 1er juillet
enquête sur les rendements 15 août	---
enquête sur les rendements 1er septembre	enquête sur les rendements 1er septembre
enquête sur les rendements[1] 15 novembre	enquête sur les rendements[1] 15 novembre

1 Cette enquête ne couvre pas les provinces maritimes et la Colombie-Britannique (excluant le district de la rivière de la Paix).

Tel que présenté dans le tableau 1, l'enquête sur les intentions d'ensemencement a lieu au début du printemps. La série des enquêtes sur les stocks de grains commence en décembre alors que les stocks sur les fermes sont à leur maximum. La deuxième enquête de ce type a lieu en mars de l'année suivante alors que la troisième se présente en juillet au moment où les stocks sont à leur niveau le plus bas. La première enquête sur les rendements, en septembre, fournit la plupart des données finales pour les Maritimes, le Québec, l'Ontario et la Colombie-Britannique (excluant le district de la rivière de la Paix) tandis que l'enquête de novembre fournit le reste des données pour le Québec et l'Ontario et l'ensemble des données finales pour les Prairies et le district de la rivière de la Paix en Colombie-Britannique.

Il est à noter que la Colombie-Britannique, sur le plan des cultures, est constituée de deux parties très différentes: le district de la rivière de la Paix qui ressemble beaucoup aux Prairies et le reste de la Colombie-Britannique qui ressemble plutôt à l'est du pays. Pour des raisons pratiques, ces deux régions ont été considérées comme deux provinces distinctes lors de la mise en place du plan d'échantillonnage.

Les sections qui suivent décrivent les différentes étapes du remaniement des enquêtes sur les cultures. Ce remaniement s'est en grande partie inspiré du plan d'échantillonnage précédent des enquêtes sur les cultures ("Panel"). Pour plus d'informations à ce sujet, voir Bélanger (1990).

2. Stratification

La stratification de la base liste est réalisée indépendamment dans chaque province canadienne. Pour les

enquêtes sur les cultures, il existe deux types de strates: les strates autoreprésentatives et les autres strates. La base de sondage n'est stratifiée qu'une fois et la stratification est la même pour chaque enquête.

2.1. Strates autoreprésentatives

2.1.1. Fermes de corporations à opérations multiples et pâturages communautaires

Certaines fermes reçoivent un traitement spécial à la collecte des données. C'est le cas des fermes de corporations à opérations multiples et des pâturages communautaires. Ainsi, par province, les pâturages communautaires sont regroupés dans une même strate autoreprésentative. Les fermes de corporations à opérations multiples forment également une strate autoreprésentative dans chaque province. Pour ces dernières, un seul contact se fait à la tête de la corporation et les données de toutes ses composantes sont alors recueillies. Dans le cadre des enquêtes sur les cultures, seules les corporations oeuvrant dans le domaine des cultures sont rejointes.

2.1.2. Fermes spécifiées par la règle de l'écart-sigma

La règle de l'écart-sigma permet d'identifier les fermes qui sont très grosses relativement à une variable donnée et qui se démarquent du reste de la population. Ces fermes, dites spécifiées, sont mises dans une strate autoreprésentative, communément appelée strate des fermes spécifiées. Cette pratique est une protection contre les valeurs aberrantes, i.e. qu'elle permet d'éviter que ces fermes viennent gonfler les estimations de façon anormale.

Une strate de fermes spécifiées a été construite dans chaque province. La variable employée est la superficie totale cultivée déclarée au Recensement de 1991. Pour les provinces des Maritimes, le Québec, l'Ontario et la Colombie-Britannique (excluant le district de la rivière de la Paix), la variable a été utilisée telle quelle en incluant la superficie en foin. Même si le foin n'est pas une variable importante dans nos enquêtes, la pratique des fermiers dans ces provinces est de faire régulièrement la rotation des cultures, y compris celle du foin. Cette superficie en foin pourrait éventuellement devenir par exemple de l'orge ou du seigle qui sont des variables importantes dans nos enquêtes. Dans la région de la Commission canadienne du blé (CCB) qui regroupe le Manitoba, la Saskatchewan, l'Alberta et le district de la rivière de la Paix en Colombie-Britannique, cette pratique de rotation n'est pas habituelle et on a donc soustrait de la superficie totale cultivée, la superficie en foin. Le foin a été défini comme étant la somme des superficies en luzerne, autre foin cultivé, autres cultures fourragères et graines de semences de plantes fourragères déclarées au Recensement de 1991. Il est important de mentionner que lors de l'identification des fermes spécifiées, les corporations à opérations multiples et les pâturages communautaires ont été exclus du processus puisqu'ils ont déjà été inclus dans des strates autoreprésentatives séparées et pourraient affecter les calculs décrits ci-dessous.

La règle de l'écart-sigma est la suivante:

Soit X_i la valeur de la variable de spécification pour l'unité i.

À partir des $X_i > 0$, on définit la médiane (MED) et l'écart-type (σ).

Soient $X_{(1)}, X_{(2)},..., X_{(N)}$ les $X_i > 0$ ordonnés.

Soit k, s'il existe, le plus petit nombre tel que $X_{(k)} > \text{MED}$ et $X_{(k)} - X_{(k-1)} > \sigma$, alors les fermes ayant $X_i \geq X_{(k)}$ sont dites spécifiées.

Pour les enquêtes sur les cultures, un total de 58 fermes sur 277 671 (0,02%) ont été classées spécifiées.

2.2. Autres strates

2.2.1. Niveaux de stratification

La province est le premier niveau de stratification. En plus de la production de bonnes estimations provinciales, la demande d'estimations fiables à des niveaux infra-provinciaux se fait de plus en plus grandissante de la part des autorités provinciales. Malheureusement, des raisons budgétaires limitent la taille de l'échantillon, et la production constante d'estimations de bonne qualité à des niveaux inférieurs devient difficile. Une des solutions envisagées était de choisir ces régions infra-provinciales d'estimation comme deuxième niveau de stratification.

Une étude a été réalisée pour la province de l'Ontario. Cette province compte 14 régions infra-provinciales d'estimation. En fixant le nombre de strates, l'étude a démontré qu'une stratification au niveau des 14 régions (i.e une strate ne peut couper deux régions), comparativement à une stratification au niveau de la province, améliorait les coefficients de variation (CV) au niveau infra-provincial, mais l'augmentation des CV provinciaux était trop importante pour justifier l'utilisation de la méthode. L'étude a tout de même fait ressortir une solution médiane: produire une stratification au niveau de régions infra-provinciales plus grandes qui sont l'union des régions infra-provinciales d'estimation. De cette façon, les CV provinciaux sont moins affectés, tandis que ceux des régions infra-provinciales d'estimation sont meilleurs que ceux de la stratification au niveau de la province seulement. Vous trouverez dans Trépanier (1993) la définition, à l'aide des divisions et subdivisions de Recensement, des régions infra-provinciales d'estimation ainsi que les regroupements effectués en vue d'obtenir des régions infra-provinciales que j'appellerai "de stratification". Ces regroupements ont été suggérés par la Division de l'agriculture qui a tenté autant que possible de joindre des régions agricoles homogènes. Cinq régions de stratification ont été formées en Ontario. Des procédures semblables ont été employées dans les autres provinces lorsque c'était nécessaire.

2.2.2. Nombre de strates

Un minimum de 4 strates par région de stratification ont été créées. La taille, en terme du nombre de fermes, et la superficie totale cultivée (en soustrayant le foin dans l'ouest) ont été calculées dans chacune des régions. Plus de strates (maximum de 10) ont été produites dans les régions où le nombre de fermes ou la superficie cultivable étaient importants.

2.2.3. Méthode de stratification

Les fermes qui ne se retrouvent pas dans une des strates autoreprésentatives décrites dans la section 2.1 ont été stratifiées dans chaque province et chaque région de stratification. Une stratification univariée a été employée avec comme variable de stratification la superficie totale cultivée moins le foin dans la région de la CCB et la superficie totale cultivée dans le reste du pays. La variable de stratification est la même que celle employée dans l'identification des fermes spécifiées afin de garder l'aspect univarié de la stratification. Les bornes des strates ont été calculées à partir d'un algorithme de Sethi (1963) lorsqu'utilisé avec la répartition optimale de l'échantillon de Neyman. Les bornes des strates et leurs tailles sont mentionnées dans Trépanier (1993).

3. Échantillon-maître

Six enquêtes au total composent la série des enquêtes sur les cultures. Afin de réduire autant que possible le fardeau des répondants, un échantillon-maître a été tiré. Celui-ci a été brisé en un certain nombre de sous-échantillons disjoints et de tailles égales dans chaque province. Le nombre et la taille des sous-échantillons ont été déterminés de façon à pouvoir utiliser pour chaque enquête un nombre de sous-échantillons dont la taille totale est proche de la taille désirée. Si un chevauchement est souhaité entre les échantillons de deux enquêtes, des sous-échantillons communs sont alors sélectionnés. Dans le cas contraire, on n'a qu'à choisir des sous-échantillons différents. L'échantillon-maître est tiré dans les "autres" strates (voir section 2.2, chapitre III). Les fermes des strates autoreprésentatives seront ajoutées aux échantillons de chaque enquête. Les sections qui

suivent expliquent les différentes étapes qui ont mené à la sélection de l'échantillon-maître et de sa partition en sous-échantillons.

3.1. Tailles des échantillons des enquêtes sur les cultures au niveau des grandes régions

Le Canada est divisé pour les besoins des enquêtes sur les cultures en trois grandes régions:

- Maritimes et Colombie-Britannique (excluant le district de la rivière de la Paix).
- Québec et Ontario.
- Région de la CCB.

Le tableau 2 montre les tailles désirées des échantillons par enquête et grande région telles que fournies par la Division de l'agriculture.

Tableau 2: Tailles désirées des échantillons de chacune des enquêtes sur les cultures par grande région.

ENQUÊTE	GRANDE RÉGION		
	Maritimes + C.-B. (excluant dist. Paix)	Québec + Ontario	CCB
Intentions-Stocks (mars)	400	2 500	5 000
Superficies (juin)	600	6 800	16 000
Stocks (juillet)	400	3 000	6 000
Rendement (septembre)	1 000	7 500	8 250
Rendement (novembre)	-	9 000[1]	16 250[2]
Stocks (décembre)	800	5 500	11 000

1 Ce nombre est approximatif car l'échantillon de novembre est en réalité constitué des fermes ayant déclaré des cultures spécialisées à l'enquête de juin ou de septembre.

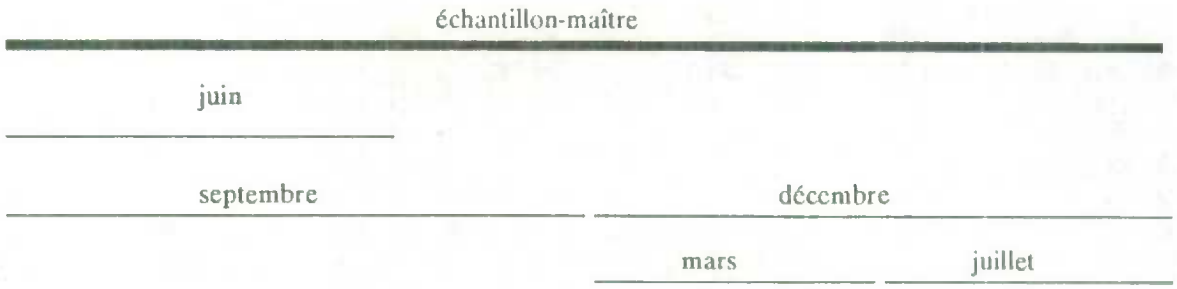
2 Ce nombre comprend un échantillon de 13 750 fermes plus les fermes ayant déclaré des cultures spécialisées à l'enquête de juin. On estime ce dernier groupe à 2 500 fermes.

La figure 1 montre la distribution de l'échantillon-maître parmi les enquêtes sur les cultures. On voit entre autres que les échantillons de l'enquête de mars et l'enquête de juillet forment l'échantillon de l'enquête de décembre. Ceci permet de produire certaines statistiques longitudinales.

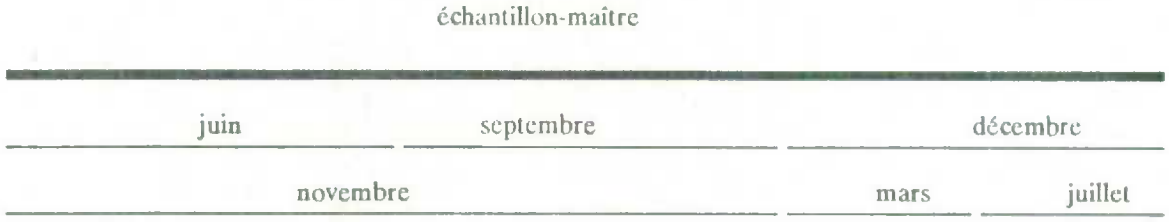
Figure 1: Distribution de l'échantillon-maître parmi les enquêtes sur les cultures.

(N. B. Ne pas tenir compte de l'échelle.)

(i) Maritimes - Colombie-Britannique (excluant le district de la rivière de la Paix)

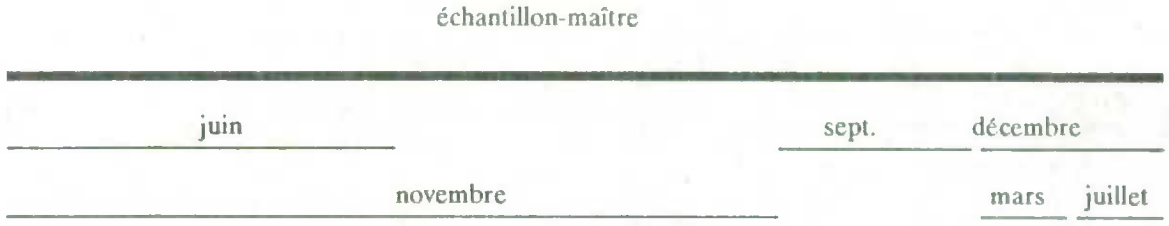


(ii) Québec et Ontario



(fermes de juin ou sept. qui ont déclaré des cultures spécialisées)

(iii) Région de la CCB



(Fermes de juin ayant déclaré des cultures spécialisées plus un échantillon supplémentaire)

3.2. Tailles des sous-échantillons

3.2.1. Taille originale

Tel que mentionné déjà, on désirait briser l'échantillon-maître en sous-échantillons de tailles égales dans une grande région R donnée. Certes il était presque impossible de trouver une taille T_R raisonnable pour un sous-échantillon telle que, pour toute enquête i donnée, $x_{iR} * T_R = n_{iR}$, où n_{iR} est la taille désirée de l'échantillon pour l'enquête i dans R et x_{iR} le nombre entier de sous-échantillons utilisés de taille T_R. Une solution triviale à cette équation est bien entendu $T_R = 1$, mais l'on cherche T_R assez grand. Nous étions par contre prêts à accepter une variation d'au plus 15% entre la taille désirée et la taille obtenue.

Ainsi on cherchait T_R tel que:

$$\frac{|x_{iR} \times T_R - n_{iR}|}{n_{iR}} \times 100 \leq 15, \forall i \in R.$$

Cet algorithme a été appliqué pour chaque grande région définie plus tôt. Plusieurs solutions étaient souvent disponibles. On a choisi celle qui avait le plus petit nombre de sous-échantillons, i.e. des sous-échantillons de grande taille.

3.2.2. Méthode du suivi limité

Au moment du remaniement, on planifiait de faire, lors de la collecte, un suivi auprès des fermes jusqu'à l'obtention d'un taux de réponse d'environ 80% pour ainsi limiter les coûts. (Le taux varie selon l'enquête et la grande région.) La taille des sous-échantillons a donc été augmentée de 25% afin de s'assurer que l'échantillon des répondants avait une taille approximativement égale à la taille désirée de l'échantillon. Puisque le biais causé par cette approche est difficilement mesurable, cette méthode a été rapidement abandonnée et il a été décidé que chaque ferme serait contactée jusqu'à 8 fois dans le but d'obtenir une réponse. Jusqu'à présent, les taux de réponse ont toujours dépassé 80%. Par conséquent, les tailles d'échantillon pourraient être révisées avant le prochain remaniement.

3.2.3. Sous-échantillons supplémentaires

Un sous-échantillon supplémentaire a été ajouté dans les Maritimes et la Colombie-Britannique (excluant le district de la rivière de la Paix), tandis que deux sous-échantillons supplémentaires ont été ajoutés dans les deux autres grandes régions. Pour le moment, ces sous-échantillons ne seront pas utilisés dans les enquêtes sur les cultures. Ils pourraient être utilisés éventuellement si un échantillon supplémentaire était demandé.

3.3. Répartition et tailles provinciales de l'échantillon-maître

Une fois que la taille des sous-échantillons par grande région est définie, on peut répartir cette taille dans les provinces qui composent ces régions. Pour ce faire, on s'est servi des tailles provinciales de l'Enquête nationale sur les fermes en 1991 afin de garder une distribution similaire au cycle précédent des enquêtes. De cette façon, la charge de travail relative par province demeure la même et de plus cette répartition s'est montrée bonne pour les estimations de culture et de bétail.

Soient n_R = taille du sous-échantillon dans la région R
n_p = taille du sous-échantillon pour la province p dans R
n_{R_ENF} = taille de l'échantillon de l'ENF 1991 dans la région R
n_{p_ENF} = taille de l'échantillon de l'ENF 1991 dans la province p dans R

Alors $n_p = n_R * n_{p_ENF} / n_{R_ENF}$ arrondi à l'unité la plus près.

La répartition obtenue du sous-échantillon est présentée dans le tableau 3 ainsi que les tailles de l'échantillon-maître.

3.4. Répartition de l'échantillon-maître dans les strates

Une répartition optimale qui minimise les CV des variables de stratification a été employée. Un programme obtenu du United States Department of Agriculture (USDA) et qui était à notre disposition a été utilisé pour

la répartition de l'échantillon-maître provincial dans les strates. Ce programme alloue une taille d'échantillon donnée de façon à minimiser une combinaison linéaire des coefficients de variation au carré des variables de répartition choisies ($\Sigma\ a_i CV_i^2$). Les CV_i dont on parle ici sont les CV théoriques calculés à partir des données de la base liste (données du Recensement de 1991) pour des variables de répartition i données. Une seule variable de répartition a été employée soit la même que pour l'identification des fermes spécifiées et la stratification. Le coefficient a_1 a été fixé à 1. Ainsi, la fonction à minimiser se ramène à CV_1^2, i.e à CV_1.

Tableau 3: Taille de l'échantillon-maître par province.

PROVINCE	ENF 1991	TAILLE D'UN SOUS-ÉCHANTILLON	NBRE DE SOUS-ÉCHANTILLONS	TAILLE ÉCHANTILLON MAÎTRE
Î.-P.-É.	487	56	6	336
Nouvelle-Écosse	571	66	6	396
N.-B.	553	64	6	384
C.-B. (sauf dist. Paix)	2 095	243	6	1 458
TOTAL	3 706	429	6	2 574
Québec	6 478	625	16	10 000
Ontario	11 193	1 080	16	17 280
TOTAL	17 671	1 705	16	27 280
Manitoba	5 079	783	20	15 660
Saskatchewan	9 609	1 482	20	29 640
Alberta	7 046	1 087	20	21 740
C.-B. (Paix)	371	57	20	1 140
TOTAL	22 105	3 409	20	68 180

3.5. Sélection de l'échantillon-maître

Afin de permettre un couplage futur avec les données fiscales, les unités de la population qui sont dans la population des dossiers fiscaux de 1991 (fermes qui ont produit un rapport d'impôt en 1991) ont premièrement été identifiées. L'échantillon-maître a été sélectionné de façon circulaire systématique dans chaque strate en opérant d'abord un tri par régions d'estimation, divisions et subdivisions du Recensement de 1991, par une variable indicatrice du lien fiscal (vaut 1 si la ferme est dans la population des dossiers fiscaux de 1991 et 0 sinon), et finalement par un nombre aléatoire préalablement attribué à chacune des fermes. Le tri effectué assure la représentativité de l'échantillon dans les domaines d'estimation et également une proportion semblable d'unités dans notre échantillon-maître liées aux dossiers fiscaux que dans notre population. Deux échantillons ont été sélectionnés. Le meilleur des deux, selon certaines statistiques étudiées telles la représentativité de l'échantillon et la performance de chacun lors de la minimisation du chevauchement (voir chapitre IV, section 3.5) a été choisi dans chaque province.

3.6. Partition de l'échantillon-maître en sous-échantillons

L'échantillon-maître sélectionné dans chaque province a été partitionné de façon aléatoire en x sous-échantillons de même taille T, où x et T sont tels qu'au tableau 3. La représentativité des strates, des régions d'estimation, des divisions et subdivisions de Recensement et des liens avec les dossiers fiscaux a été assurée. Diverses statistiques ont été calculées afin de s'assurer que chacun des sous-échantillons représente bien la population. Les tailles des échantillons provinciaux des enquêtes sont présentées dans le tableau 4.

Tableau 4: Taille des échantillons provinciaux des enquêtes sur les cultures.

PROV.	TAILLES DES ÉCHANTILLONS PAR ENQUÊTE[1] (nombre de sous-échantillons utilisés)					
	MARS	JUIN	JUILLET	SEPT.	NOV.	DÉC.
Î.-P.-É	56 (1)	112 (2)	56 (1)	168 (3)	-	112 (2)
N.-É.	66 (1)	132 (2)	66 (1)	198 (3)	-	132 (2)
N.-B.	64 (1)	128 (2)	64 (1)	192 (3)	-	128 (2)
C.-B. (sauf Paix)	243 (1)	486 (2)	243 (1)	729 (3)	-	486 (2)
Québec	1 250 (2)	3 125 (5)	1 250 (2)	3 125 (5)	Fermes de juin ou sept. ayant déclaré cultures spécialisées	2 500 (4)
Ontario	2 160 (2)	5 400 (5)	2 160 (2)	5 400 (5)		4 320 (4)
Manitoba	1 566 (2)	4 698 (6)	1 566 (2)	2 349 (3)	3 915[2] (5)	3 132 (4)
Sask.	2 964 (2)	8 892 (6)	2 964 (2)	4 446 (3)	7 410[2] (5)	5 928 (4)
Alberta	2 174 (2)	6 522 (6)	2 174 (2)	3 261 (3)	5 435[2] (5)	4 348 (4)
dist. Paix	114 (2)	342 (6)	114 (2)	171 (3)	285[2] (5)	228 (4)

1 Ces tailles n'incluent pas les fermes spécifiées, les pâturages communautaires et les corporations à opérations multiples.

2 On ajoute à ce nombre les fermes ayant déclaré des cultures spécialisées en juin.

IV.ENQUÊTES SUR LE BÉTAIL

1. Aperçu

L'Enquête nationale sur les fermes (ENF) avait jusqu'en 1992 le mandat de faire la collecte des données sur le bétail, sur les cultures et sur les revenus et dépenses des fermiers. La collecte des informations de bétail se faisait en deux temps. En juin, les fermes échantillonnées étaient contactées une première fois pour leurs cultures. Quelques questions sur le bétail étaient également posées. En juillet, les fermes identifiées comme faisant l'élevage du bétail lors du premier contact étaient recontactées et questionnées plus en détails sur le bétail. Un sous-échantillon de l'échantillon de l'ENF de juillet était à nouveau contacté en janvier pour l'Enquête de janvier sur les fermes (EJF). Les fermiers devaient alors répondre à une série de questions portant principalement sur le bétail.

Le présent remaniement propose de garder les deux principales enquêtes décrites ci-dessus. L'Enquête de juillet sur le bétail aura lieu en juin (avec date de référence au 1er juillet). L'Enquête de janvier sur le bétail remplacera l'ancienne EJF et encore une fois son échantillon sera un sous-échantillon de l'échantillon de juillet de l'année précédente. La différence la plus importante dans ce remaniement demeure les caractéristiques mêmes du plan d'échantillonnage. Celui-ci (entre autres dans la stratification) sera développé à partir de variables de bétail uniquement plutôt qu'une multitude de variables agricoles comme c'était le cas avec l'ENF. La base de sondage sera la même que celle des enquêtes sur les cultures.

2. Stratification

2.1. Strates autoreprésentatives

2.1.1. Fermes de corporations à opérations multiples et pâturages communautaires

De façon identique à ce qui a été fait pour les enquêtes sur les cultures, toutes les fermes d'une même province faisant partie d'une corporation à opérations multiples ont été mises dans une même strate autoreprésentative. Parmi ces fermes, seules les corporations à opérations multiples oeuvrant dans le domaine du bétail seront contactées. Les pâturages communautaires d'une même province ont aussi été regroupés dans une même strate autoreprésentative.

2.1.2. Fermes spécifiées par la règle de l'écart-sigma

La règle de l'écart-sigma a une fois de plus été employée afin d'identifier certaines fermes importantes pour les variables suivantes: vaches de boucherie, vaches laitières, truies, total en bovins, total en porcs et total en moutons. Plus exactement, la règle de l'écart-sigma a été appliquée sur l'ensemble des fermes d'une même province (en excluant les pâturages communautaires et les corporations à opérations multiples) pour chacune des six variables précédentes. Les fermes ayant été identifiées pour au moins une variable sont regroupées dans une strate appelée strate des fermes spécifiées. Pour les enquêtes sur le bétail, les strates de fermes spécifiées comptent 223 fermes, dont 8 seulement chevauchent les strates de fermes spécifiées pour les enquêtes sur les cultures. Ce nombre (8) est très faible et semble montrer qu'une ferme très importante pour les cultures ne l'est pas nécessairement pour le bétail.

2.2. Strate de fermes sans bétail

Selon les données du Recensement, plusieurs fermes sur la base de sondage liste ne font pas du tout l'élevage du bétail en termes de bovins, porcs et moutons. Par contre, ces fermes n'ont pas été exclues de la population cible car elles peuvent commencer l'élevage du bétail au cours de la durée du nouveau plan d'échantillonnage. De plus, une ferme peut ne pas avoir été bien classée au recensement. Pour ces raisons, ces fermes dites sans bétail (i.e sans bovins, porcs et moutons) ont été regroupées dans une seule strate et ce, pour chaque province. Cette strate, on le verra plus loin, sera échantillonnée à un faible taux.

2.3. Autres strates

Dans chaque province, toutes les fermes n'étant pas déjà incluses dans une des strates précédentes (autoreprésentatives ou sans bétail) ont été stratifiées selon la méthode décrite dans les sections qui suivent.

2.3.1. Variables de stratification

La stratification est de type multivariée et les variables employées sont les mêmes que pour l'obtention des fermes spécifiées (voir section 2.1.2) à l'exception qu'elles sont standardisées (moyenne = 0, variance = 1). Ceci permet d'éviter la dominance d'une grosse variable (ex.: total en bovins) sur une variable de moindre taille. Avant d'en arriver à cette transformation des variables, d'autres méthodes ont été testées. Entre autres, nous avons tenté d'utiliser les variables originales non-transformées, les composantes principales des six variables de bétail originales, ou les variables transformées à l'aide de la division de la variable originale par la moyenne des contributeurs non-nuls à cette variable. Aucune de ces transformations ne s'est avérée meilleure que la standardisation.

2.3.2. Procédures de stratification

Les fermes d'une même province ont été stratifiées à l'aide des procédures multivariées FASTCLUS et CLUSTER du progiciel SAS. Pour plus de détails sur ces procédures, voir SAS Institute Inc. (1985). Le nombre de fermes à stratifier étant trop élevé, nous avons dû appliquer d'abord la procédure FASTCLUS. Celle-ci a permis un regroupement initial des fermes en 150 grappes en un maximum de 10 itérations. Nous avons ensuite appliqué la procédure CLUSTER aux 150 grappes, qui sont en fait devenues 150 unités à stratifier, pour former

une stratification hiérarchique allant de 150 strates à 1 strate. La méthode de la variance minimum de Ward (1963) a servi pour cette stratification. Les variables de stratification dans CLUSTER ont légèrement été modifiées. Ce sont les moyennes des variables de stratification originales (variables standardisées) pour une grappe donnée. Plus précisément, soient $i = 1,2,...,N_c$ les unités de la grappe c et $c = 1,...N$, les grappes données par FASTCLUS. Soit X la variable de stratification originale (variable de bétail standardisée, moyenne = 0, variance = 1). Alors la nouvelle variable de stratification utilisée par la procédure CLUSTER est $\bar{X}$ où

$$\bar{X}_c = \frac{1}{N_c}\sum_{i=1}^{N_c} X_i, c = 1,...,N.$$

2.3.3. Nombre de strates par province

Le choix du nombre de strates est basé sur deux objectifs précis mais opposés: réduire les coefficients de variation (en augmentant le nombre de strates) des variables importantes et garder à long terme une stratification robuste (en diminuant le nombre de strates) qui réduira le nombre d'unités qui sauteront d'une strate à l'autre sur une période de 5 ans (durée d'existence du nouveau plan d'échantillonnage). Il est difficile d'évaluer présentement la robustesse d'une stratification tant que l'enquête n'est pas menée. Cependant les CV ont été calculés pour plusieurs nombres de strates en fixant la taille provinciale de l'échantillon et en la répartissant dans les strates selon la méthode décrite à la section 3.3. Cet exercice a permis de déterminer le nombre minimal de strates qui permettait d'atteindre une précision acceptable.

La procédure TREE de SAS a permis d'extraire le nombre désiré de strates de la stratification hiérarchique produite par CLUSTER. Puisque la différence entre h et h + 1 strates est minime, nous avons extrait différentes stratifications à intervalles de 2 ou 3 strates. Les données du Recensement ont été utilisées pour calculer les coefficients de variation théoriques à l'aide de la formule suivante:

Soit X la variable d'intérêt et x_i la valeur de cette variable pour l'unité i sur la base de sondage (i.e au Recensement de 1991). Soient également

H = nombre de strates dans une province donnée
N_h = nombre d'unités dans la strate h
n_h = nombre d'unités échantillonnées dans la strate h

$$CV(X) = \frac{\sqrt{\sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h(N_h - 1)} \sum_{i=1}^{N_h} (x_i - \bar{x})^2}}{\sum_{h=1}^{H}\sum_{i=1}^{N_h} x_i}.$$

Le nombre total de strates (incluant les strates autoreprésentatives et la strate sans bétail) varie de 15 à 46 d'une province à l'autre.

3. Échantillon-bidon de juillet 1992

3.1. Aperçu

La première enquête sur le bétail basée sur le nouveau plan d'échantillonnage a eu lieu en janvier 1993. Cela

posait un problème au moment du remaniement puisque l'échantillon de janvier est un sous-échantillon de celui du mois de juillet de l'année précédente. En guise de solution, un échantillon-bidon a été tiré pour juillet 92 même si l'enquête n'a effectivement pas eu lieu (juillet 92 a été couvert par l'ENF). Cet échantillon n'a pas servi, mais on a pu en extraire un sous-échantillon pour l'enquête de janvier 93. Une rotation des deux échantillons est prévue pour les années subséquentes. Pour juillet 93, l'échantillon-bidon de juillet 92 modifié par la rotation sera utilisé.

3.2. Répartition et tailles provinciales de l'échantillon

Le budget permettait un échantillon d'environ 29 750 unités pour juillet. Pour déterminer la proportion de l'échantillon total attribuée à une province, on s'est inspirée de ce qu'elle était pour l'ENF de 1991 (voir Chapitre III, section 3.3). Les tailles des échantillons provinciaux sont présentées dans le tableau 5.

Tableau 5: Tailles provinciales de l'échantillon-bidon de bétail de juillet 1992 en excluant les corporations à opérations multiples.

PROVINCE	Échantillon-bidon de juillet 1992
Île-du-Prince-Édouard	335
Nouvelle-Écosse	393
Nouveau-Brunswick	381
Québec	4 420
Ontario	6 801
Manitoba	3 504
Saskatchewan	5 583
Alberta	6 631
Colombie-Britannique	1 703
TOTAL	29 751

3.3. Répartition de l'échantillon dans les strates

La fraction de sondage pour la strate des fermes sans bétail a été fixée à 1/60. En soustrayant de la taille présentée au tableau 5, le nombre de fermes spécifiées, de pâturages communautaires et de fermes échantillonnées dans la strate des fermes sans bétail, on obtient l'échantillon qui doit être réparti dans les autres strates d'une province. Les corporations à opérations multiples ne sont pas incluses dans nos calculs, car leur nombre est minime et encore indéterminé. La répartition de l'échantillon dans les autres strates s'est faite exactement comme pour les enquêtes sur les cultures (voir chapitre III, section 3.4) selon les variables utilisées dans l'identification des fermes spécifiées. Les coefficients a_1, a_2, ..., a_6 ont tous été fixés à 1 puisque toutes les variables étaient jugées d'importance égale.

3.4. Sélection de l'échantillon-bidon de juillet 1992

L'échantillon-bidon de juillet 92 a été sélectionné de façon circulaire systématique. Cette fois avant de sélectionner le nombre approprié d'unités dans une strate, les unités ont d'abord été triées par divisions et

subdivisions de Recensement, par variable indicatrice du lien fiscal et ensuite par un nombre généré aléatoirement. Deux échantillons ont encore une fois été sélectionnés. Le meilleur des deux, selon certaines statistiques étudiées telles la représentativité de l'échantillon et la performance de chacun lors de la minimisation du chevauchement (voir chapitre IV, section 3.5), a été choisi pour chaque province.

3.5. Minimisation du chevauchement de l'échantillon-bidon de juillet 1992 avec l'échantillon-maître des enquêtes sur les cultures.

Afin de répartir le fardeau de réponse, le chevauchement entre l'échantillon-bidon de juillet 1992 et l'échantillon-maître des enquêtes sur les cultures a été minimisé. Pour ce faire, nous avons adapté pour nos besoins la méthode de Kish et Scott (1971) pour probabilités initiales de sélection égales. Dans chaque intersection d'une strate des enquêtes sur le bétail et d'une strate des enquêtes sur les cultures, les unités faisant partie des deux échantillons sont identifiées. Pour l'échantillon de bétail, ces unités sont remplacées aléatoirement par des unités disponibles dans l'intersection des strates. Ces unités qui faisaient originalement partie des deux échantillons et qui ont été remplacées demeurent dans l'échantillon-maître de cultures. Vous trouverez dans les prochaines lignes une description plus détaillée de la méthode employée. Le lecteur intéressé pourra trouver dans Perry, Burt et Iwig (1993) deux méthodes de réduction du chevauchement lorsque deux enquêtes ou plus sont impliquées. La première méthode réduit le chevauchement par l'entremise de la programmation linéaire en nombres entiers tandis que la deuxième est la méthode ci-dessous, généralisée au cas de k enquêtes ($k \geq 2$).

Soient $U_{c.}$ et $U_{.b}$ les strates de cultures et de bétail respectivement. Soient également $U_{cb} = U_{c.} \cap U_{.b}$ et

e_{cb} = unités à la fois dans l'échantillon de cultures et de bétail
n_{cb} = nombre d'unités dans e_{cb}
U'_{cb} = unités de U_{cb} qui ne sont dans aucun échantillon
N'_{cb} = nombre d'unités dans U'_{cb}

Deux actions sont possibles:

(1) Si $n_{cb} \leq N'_{cb}$, n_{cb} unités sont sélectionnées par échantillonnage aléatoire simple dans U'_{cb}. Les unités sélectionnées remplacent désormais les unités de e_{cb} dans l'échantillon de bétail.
(2) Si $n_{cb} > N'_{cb}$, N'_{cb} unités sont sélectionnées par échantillonnage aléatoire simple dans e_{cb}. Les unités sélectionnées sont remplacées par les unités de U'_{cb} dans l'échantillon de bétail.

Lors de la minimisation du chevauchement à l'intérieur d'une intersection U_{cb} donnée, une unité n'était pas nécessairement remplacée par une unité ayant les mêmes caractéristiques telles la division et subdivision de Recensement, ou encore par une unité liée à un dossier fiscal si l'unité originale l'était. Deux échantillons ont été sélectionnés pour bétail et pour cultures. La minimisation du chevauchement a été réalisée sur chaque combinaison d'un échantillon de cultures et d'un échantillon de bétail dans chaque province. La pire combinaison a donné un chevauchement final de 4799 unités suite à l'application de la méthode de minimisation du chevauchement. Selon la performance de chaque combinaison et des statistiques de chacun des échantillons, un échantillon a été choisi pour bétail et cultures. À l'aide de la méthode décrite plus haut, le chevauchement entre l'échantillon-bidon de bétail et l'échantillon-maître de cultures a pu être réduit de 58% pour résulter en un chevauchement final de 4631 unités.

4. Échantillon de janvier 1993

4.1. Répartition et tailles provinciales de l'échantillon

L'échantillon pour l'enquête de bétail de janvier est un sous-échantillon de l'échantillon de juillet. Le budget permettait une taille d'échantillon pour janvier de 13 500 unités. En s'inspirant de la répartition de l'EJF de 1992, ces 13 500 unités ont été réparties dans les provinces. Dans les Maritimes, l'échantillon de janvier est le

même que celui de juillet. Les tailles finales sont présentées dans le tableau 6.

4.2. Répartition de l'échantillon dans les strates

La stratification pour l'Enquête sur le bétail de janvier est la même que pour l'Enquête sur le bétail de juillet. L'échantillon de juillet dans les strates autoreprésentatives et dans la strate des fermes sans bétail est gardé en janvier. Donc une certaine partie des tailles d'échantillons au tableau 6 est accaparée par ces strates. La taille provinciale restante de l'échantillon de janvier a été répartie indépendamment de celle de juillet. Il est évident que pour une strate donnée, la taille allouée à cette strate est plus grande pour juillet que pour janvier puisque l'échantillon total est au moins d'égale importance. Le programme du USDA a encore une fois été employé pour la répartition. Les variables de répartition sont les mêmes que pour juillet.

4.3. Sélection de l'échantillon

Le nombre alloué d'unités à chaque strate a été sélectionné parmi les unités échantillonnées en juillet pour la même strate. À l'intérieur des strates, les unités ont été triées par divisions et subdivisions de Recensement, variable indicatrice du lien fiscal et un nombre généré aléatoirement pour chaque unité. La sélection s'est ensuite faite de façon circulaire systématique.

Tableau 6: Tailles des échantillons provinciaux pour l'Enquête de janvier sur le bétail, excluant les corporations à opérations multiples.

PROVINCE	TAILLE DE L'ÉCHANTILLON PROVINCIAL
Île-du-Prince-Édouard	335
Nouvelle-Écosse	393
Nouveau-Brunswick	381
Québec	2 297
Ontario	2 718
Manitoba	1 349
Saskatchewan	2 385
Alberta	2 656
Colombie-Britannique	988
TOTAL	13 502

V. CONCLUSION

Les nouvelles enquêtes sur les cultures et sur le bétail emploient également une base aréolaire. Cet aspect n'a pas du tout été traité dans ce rapport. Un document de travail séparé présentant le remaniement de la base aréolaire devrait être disponible sous peu. De plus, nous devrions être en mesure de qualifier et de quantifier au cours des prochaines années l'impact du nouveau remaniement sur les enquêtes sur les cultures et sur le bétail. Certes le plan d'échantillonnage a été défini spécifiquement pour chacun des types d'enquêtes, contrairement à l'ENF qui devait tenir compte à la fois de trois types de données agricoles. Ce remaniement, plus adapté aux besoins de chacun, laisse espérer une meilleure efficacité à bien des points de vue, mais le prix

à payer est l'augmentation du fardeau du répondant et la perte des liens entre les variables de cultures, de bétail et financières.

VI. BIBLIOGRAPHIE

Bélanger, Y. (1990), "Le remaniement des enquêtes sur les cultures à Statistique Canada", Cahier de travail de la Direction de la méthodologie de Statistique Canada, numéro BSMD-90-014E/F.

Julien, C. et Maranda, F. (1990), "Le plan de sondage de l'Enquête nationale sur les fermes de 1988", *Techniques d'enquêtes*, 16, 117-129

Kish, L. et Scott, A. (1971), "Retaining Units after Changing Strata and Probabilities", *Journal of the American Statistical Association*, 66, 461-470.

Perry, C., Burt, J. et Iwig, B. (1993), "Methods of Selecting Samples in Multiple Surveys to Reduce Respondent Burden", *Proceedings of the International Conference on Establishment Surveys, American Statistical Association,* à venir.

Sas Institute Inc. (1985), *SAS User's Guide: Statistics,* Version 5, Cary, North Carolina: SAS Institute Inc.

Sethi, V. K. (1963), "A Note on Optimum Stratification of Populations for Estimating the Population Means", *Australian Journal of Statistics*, 5, 20-33.

Trépanier, J. (1993), "Remaniement des enquêtes agricoles sur les commodités: Base liste", Rapport technique de Statistique Canada.

Ward, J. H. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236-244.