SSC Annual Meeting, June 2005
Proceedings of the Survey Methods Section

IMPROVING THE QUALITY OF THE ESTIMATES OF SPECIAL CROPS

Laurie Reedman[1]

ABSTRACT

The series of Crops Surveys collect information from Canadian crop producers six times each year. These data are used to estimate the acreage and yield of all principal crops, such as wheat, corn, barley and oats. Also, estimates of many "special" crops are produced such as lentils, dry peas and beans, and canary seed. These special crops are only grown by a small percentage of Canadian producers, but they play a significant role in the agriculture industry. This paper explores the use of a composite estimator to improve the accuracy and precision of estimates of certain special crops while not increasing the regular sample size.

KEY WORDS: Accuracy; Composite estimator; Precision, Rare characteristics, Special crops.

RÉSUMÉ

La série d'enquêtes sur les cultures recueille de l'information auprès des exploitants agricoles canadiens six fois par année. Ces données sont utilisées pour l'estimation de la superficie et du rendement de tous les grandes cultures, comme le blé, le maïs, l'orge et l'avoine. On produit également des estimations pour plusieurs « cultures spéciales » dont les lentilles, les pois secs, les haricots secs, et les graines de l'alpiste des Canaries. Ce sont certes là des cultures menées par un petit pourcentage des exploitants agricoles canadiens, mais elles jouent néanmoins un rôle significatif dans l'industrie de l'agriculture. Dans cet article, on examine la possibilité de recourir à un estimateur composite pour améliorer l'exactitude et la précision des estimations de certaines cultures spéciales tout en conservant la même taille d'échantillon.

MOTS CLÉS : Cultures spéciales; estimateur composite; l'exactitude; précision; characteristiques rares.

1. INTRODUCTION

As described by Morrison (2001), the Crops panel of surveys is a series of six probability surveys conducted every year. The Crops Surveys collect and disseminate data on seeding intentions, seeded and harvested area, yield, production and stocks for field crops in Canada. The target population includes agricultural operations producing agricultural products for sale.

The March Crops Survey provides estimates of what producers are intending to plant in their fields. The June Crops Survey measures seeded area, and the September Crops Survey gives an early indication of crop production. The July Crops Survey measures stocks at their lowest point, for comparison with the December Crops Survey which measures stocks at their highest. The November Crops Survey is the "flagship" occasion, from which are produced estimates of harvested area, yield, and production.

The population of crops farms in Canada is approximately 190,000 farms. Samples range in size from 12,000 in March to 32,000 in November. Simple random sampling with rotation is used and a response rate of 80% is usually achieved.

Several estimation issues are raised by the Crops surveys. Some of these have already been addressed by Reedman (2004). This paper addresses other issues related to special crops. Section 2 describes the estimation process for principal crops. The special crops producers are described in Section 3, and Section 4 proposes the use of a composite estimator along with various design initiatives in order to improve special crops estimates.

---

[1] Laurie Reedman, Statistics Canada, 120 Parkdale Avenue, Ottawa, ON, K0A 1T6, Laurie.Reedman@statcan.ca

2. ESTIMATING PRINCIPAL CROPS

The sample is designed to produce estimates of the principal crops wheat, oats, barley, hay and canola with a coefficient of variation of 2% at the province level. A list frame is used, stratified by province, agricultural region within province, and acreage class. Simple random sampling is used within strata. Sample rotation is used to ease response burden. Overlap with other Agriculture surveys is managed through the use of permanent random numbers. As proposed by Cochran (1977), totals are estimated by

$$\hat{Y} = \sum_{h} \sum_{i \in h} y_i w_h$$

where h is the stratum, i is the index for units within strata, y is the observed commodity for unit i, and w_h is the weight within stratum h. The coefficient of variation is estimated by

$$\frac{\sqrt{\text{var}(\hat{Y})}}{\hat{Y}}.$$

3. SPECIAL CROPS PRODUCERS

The sample is used to produce estimates for approximately one hundred different field crops, some that are less commonly grown than the principal crops from which the sample size is determined. In particular, there are forty crops that are termed “special crops”. These include eight different kinds of dry beans, two kinds of chick peas, three kinds of lentils, as well as caraway, linola, and triticale. See Appendix A for a complete list of the special crops. These are of particular interest because they play a significant role in the agriculture industry. These crops are staples in the diets of many African and Asian cultures, and hence have significant market value. The Canadian Special Crops Association has lobbied Agriculture and Agri-Food Canada to learn more about the production of these crops in Canada. Statistics Canada in turn uses the Crops Survey for this purpose.

The producers of special crops form a rare sub-population, making up just 12% of the entire population of crops producers. They are also a mobile population. Approximately half of the producers who reported some special crops on the 1996 Census of Agriculture did not report any special crops on the 2001 Census of Agriculture, and roughly half of those reporting special crops on the 2001 Census of Agriculture did not report any previously. Some special crops such as canary seed require specialized equipment. If a producer has invested in specialized equipment it is likely that he or she will continue to grow that particular crop. However, in other cases, a producer may try out a special crop in hopes that it will provide good revenue. In this way, a producer may report special crops in one year and not in another year. It is also beneficial to rotate some crops with others to replenish nutrients in the soil, producing the illusion for survey takers that the farm is moving in and out of the special crops target population. There are approximately 22,000 special crops producers in the population, and most of those farms also produce some principal crops.

The estimators as described above do not produce good quality estimates for the special crops. This is because the sample size is not large enough, and the stratification does not create homogenous groups of special crops producers. What is sought after is a method to improve the quality of the estimates of special crops, without increasing the overall sample size, and without impacting the quality of the estimates of the principal crops.

4. ESTIMATING SPECIAL CROPS

4.1 Composite Estimation

Estimates of special crops are produced from all six of the Crops Survey occasions, however this proposal focuses on the November survey because it has the largest sample, producing the best quality estimates to date, and also because it covers the widest range of estimated variables (harvested area, yield and production of more than one hundred commodities). The proposal is to re-organize the sample into two parts. One part would remain as a simple random sample from the full population, using the current stratification, so as to produce the required estimates of the principal crops. The other part of the sample would be a stratified simple random sample from the same population but stratified such that special crops producers are isolated into homogenous groups. A composite estimator as described by Sigman and Monsour (1995) would be used to incorporate both parts of the sample in the estimates of special crops.

With this approach, there would be two estimates for the same commodities, representing the same population, but coming from two independent samples. A composite estimator could simply be the average of the two estimates, or a weighted average, where more emphasis is put on one sample than the other. One option would be to weight the estimates according to their contribution to the combined sample size.

More specifically, if we let $\hat{Y}_S$ be the estimate of a given special crop from the special crops sample, and let $\hat{Y}_R$ be the estimate of the same special crop from the regular sample, then the composite estimator $\hat{Y}$ can be expressed as $\hat{Y} = \alpha\hat{Y}_S + (1-\alpha)\hat{Y}_R$ where α is between 0 and 1. The variances of $\hat{Y}_S$ and $\hat{Y}_R$ can also be combined, as described by Kalbfleish (1985):

$$\text{var}(\hat{Y}) = \text{var}(\alpha\hat{Y}_S + (1-\alpha)\hat{Y}_R) = \alpha^2\,\text{var}(\hat{Y}_S) + (1-\alpha)^2\,\text{var}(\hat{Y}_R) + 2\alpha(1-\alpha)\,\text{cov}(\hat{Y}_S, \hat{Y}_R)$$

The two samples are independent, therefore the covariance of $\hat{Y}_S$ and $\hat{Y}_R$ is zero.

The coefficient of variation of the composite estimator is the square root of the variance of the composite estimator divided by the composite estimator itself. The coefficient of variation of the composite estimator is usually less than the coefficient of variation of either of its components.

We would expect the sample stratified to isolate special crops producers to produce a better quality estimator for special crops, but possibly a worse quality estimator for principal crops. However, it is desirable to use all the data pertaining to principal crops that would be collected from the special crops producers. This might necessitate the use of different values of α for principal and special crops. An α value could be chosen to minimize the coefficient of variation, producing an estimator with good precision. Alternatively, an α value could be chosen to minimize bias, producing an estimator with good accuracy. The risk of bias lies in the special crops stratification. Ideally this stratification would produce some strata with few farms but a high concentration of special crops producers, and other strata with many farms and few special crops producers. If sample is drawn from the strata where there are few special crops producers, these units will have high sampling weights and outlier observations could result. If on the other hand no sample is drawn from those strata, some special crops producers would likely be missed, resulting in under coverage bias.

4.2 Challenges

The success of a composite estimator at improving the quality of the special crops estimates depends on the stratification method. An indicator for having reported special crops on the Census of Agriculture or a previous Crops survey has been investigated, but was not satisfactory. Neither the 2001 Census of Agriculture nor Crops Surveys prior to September 2004 asked about the production of all of the special crops, consequently too many farms would be missed by a flag based on these sources. The use of a geographic indicator was also explored, but without promising results. Even in regions where there is a high concentration of special crop production, there are also many farms with no special crops, and the converse is also true, that is, there are many special crops producers located outside of the regions of high concentration. Other auxiliary variables have been tested (for example, the use of irrigation or the production of known “companion” crops to some special crops) but these also were not satisfactory at isolating special crops producers.

4.3 Future Work

Future work on this project will include an exploration of the data gathered in the 2006 Census of Agriculture. In the Census, all agriculture producers in Canada will be asked their acreage of many (but unfortunately not all) of the special crops. This will allow for the identification of many of the special crops producers. The new Census will also afford the opportunity to redesign the Crops Survey. It might be possible to increase the rate of sample rotation such that most or all of the population can be surveyed in the five year inter-censal period. This would provide a rich source of data, allowing for the identification of all farms producing any one of the forty different special crops. It might then be possible to determine a profile of special crops producers, which would translate into a set of auxiliary variables that would serve as better special crops stratification variables.

Another avenue for future work lies in the composite estimator itself. Stratification variables can be simulated so that work can proceed in determining the optimal α values under various constraints.

5. CONCLUSIONS

Given that the special crops producers are a rare and mobile sub-population within the regular survey frame, the estimation method used for the principal crops does not produce good quality estimates of the special crops. Improvements to the quality of the estimates of special crops cannot be made through methods that would jeopardize the quality of the estimates of principal crops, nor can the overall sample size be increased. A composite estimation method is proposed, with one sample drawn from the population stratified for principal crop producers, and the other sample drawn from the same population but stratified to isolate the special crop producers. The relative contribution of each estimator can vary depending on accuracy and precision requirements. The success of the composite estimator depends on the availability of good stratification information. More work is required to identify appropriate stratification variables.

APPENDIX A

The special crops are:

lentils small green
lentils large green
lentils red
lentils other
dry peas yellow
dry peas green
dry peas other
dry white beans
mustard other
mustard yellow
mustard oriental
mustard brown
sunflower
fababeans
triticale
safflower
caraway
coriander
borage
linola
buckwheat
corn for grain
fodder corn
potatoes
soybeans
canary seed hairless
canary seed regular
chick peas desi
chick peas kabuli
chick peas other
dry beans black
dry beans cranberry
sugar beets
dry beans dark red kidney
dry beans great northern
dry beans light red kidney
dry beans pinto
dry beans small red
dry beans other
flaxseed

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques* (3rd. ed.). New York: John Wiley.

Kalbfleisch, J.G. (1985). *Probability and Statistical Inference Volume 1: Probability* (2nd. Ed.). New York Berlin Heidelberg Tokyo: Springer-Verlag.

Morrison, R. (2003). *2001 Agriculture Redesign*, BSMD Statistics Canada working paper.

Reedman, L. (2004). *Methods To Improve the Level Estimates in the Crops Surveys*, BSMD Statistics Canada working paper.

Sigman, R. and Monsour, N. (1995). *Selecting Samples from List Frames of Businesses*. Business Survey Methods, edited by Cox, Binder, Chinnappa, Christianson, Colledge, Kott. John Wiley.