The Statistical Analysis of the Economic Impact of Hog Production in Manitoba Communities: The Pembina Valley Region

BY

Lesley Crisostomo

A Practicum submitted to the Faculty of Graduate Studies in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Statistics

University of Manitoba

Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

The Statistical Analysis of the Economic Impact of Hog Production in Manitoba Communities: The Pembina Valley Region

BY

Lesley Crisostomo

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

Master of Science

Abstract

This paper investigates the sampling techniques and estimation procedure using survey information. A problem in survey data are the nonresponses. In this paper, I describe how to use single imputation to obtain estimates and their variances when some individuals have missing data. The procedures used are: proportion, ratio, and regression estimation.

Acknowledgements

There are few people that I would like to acknowledge that have contributed in the process of this research project. Dr. Jim MacMillan was the coordinator of the project. He proposed the study and founded a group of producers, provincial government and researchers to work on the project. Slyvio Sabourin, Qin Chen, Dr. Bruce Johnston, and I, Lesley Crisostomo, were asked to take part in his study. Dr. Bruce Johnston and Dr. Jim MacMillan defined the population and sampling frame. Slyvio Sabourin distributed and collected the questionnaire for the producer in the municipalities in the Pembina Valley region. Qin Chen gathered the information from Slyvio Sabourin and organized the information into matrix form. I used this information to determine the properties of hog production. Qin Chen used the estimation for the input-output model. Dr. Jim MacMillan, Dr. Bruce Johnston, Qin Chen, and I reviewed the results of the information from the input-output model. Dr. Jim MacMillan reported the estimates of production and the input-output analysis for the Agricultural Committee of the Pembina Valley Regional Development Corporation.

I would like to further thank my supervisor, Dr. Bruce Johnston for giving me the opportunity to work on this project. With his guidance, support and suggestions, it gave me determination to complete this report. I would also like to thank Dr. MacMillan and Qin Chen for the opportunity to work with them on this project. Their answers to my questions provided me useful information. Dr. Mount, a member of my committee, thank you for taking the time to help me. I am

extremely grateful to my family and friends for their support, encouragement and for believing in me. Stella Leung and Tony Wu, you two have been the greatest classmates and have inspired me always. To my boyfriend, Jeremy, thank you for being there and supporting me through my master's degree.

Table of Contents

List of Tables

Chapter 1

Introduction

The Agricultural Committee of the Pembina Valley Regional Development Corporation requested research on the regional economic impacts of hog production. The expansion of hog production in the province of Manitoba motivated this study to take place. This study is one of the first of this kind. The total economic impact of hog production of Manitoba has not previously been quantified. Therefore the outcome of this project will be important to policy makers. One can expect that the policy will change as hog production expands. The objective of this practicum paper is to determine a frame and a procedure to estimate the information needed to find the economic impact of hog production. This information will be important to the Agricultural community.

1.1 Research objectives

The main focus is on the sampling and estimation analysis of hog production in the Pembina Valley Region. The sample design for the small producers was to survey 4 producers from each municipality making the total small producers surveyed 32. All of the remaining medium and large producers were to be contacted. Therefore, our sample consists of 97 producers (32 small producers, 57 medium producers and 8 large producers). A later chapter will include a more detailed illustration of the sample procedure. The estimation analysis deals with

components of hog production. Some components are labour, operation cost, revenue, number of hogs sold, and exports. For simplicity, the examples in this paper will use labour as the component of interest. There were three different types of estimation conducted in this study, i.e. proportion, ratio and regression estimation.

In the sampling process the issue of nonresponse became of some concern. Specifically less than 50% of medium hog producers responded. From the large producers, 7 out of the 8 responded. Out of the 32 sample of the small hog producers 21 replied. There are various techniques to accommodate the problem of nonresponse, including the use of imputation. This procedure estimates missing data using information at hand. An assumption was made that nonrespondents' answers could be predicted based on the patterns of the respondents' characteristics and answers. Specifically, the animal units and producer's type of operation were used to predict their response. Animal units are the number of animals required to excrete a total of 73 kg of nitrogen in one year[1]. Producer's type of operation consists of 6 different types of operation (ex. Farrow to Finish).

1.2 Background Information

Manitoba is the third largest hog-producing province, after Ontario and Quebec. Manitoba produced about 23% of the Canadian total of hogs and 4% of the North

[1] Economic impacts of hog production in Manitoba communities/regions: the Pembina Valley Region (J. MacMillan et Al., 2004)

American total[2]. Hog production in Manitoba is increasing. In 2000, Manitoba's 1, 430 hog operations produced 5.35 million hogs, an increase of 12.4% from 1999[3].

Exports to the United States are the major reason for growth in hog production in Canada. Hog and pork exports to the United States depend on current tariff negotiations. Since 1997 the demand for pork has grown rapidly, especially in Asia. In some studies it shows that Canada is the only major exporting country, which has had steady growth in hog production over the past several years (Alexiou et al, 1999). Manitoba is the largest hog-exporting province in Canada. About 200 million dollars worth of hogs were shipped to the US and Mexico in 2000 compared to 165 million dollars worth in 1999[4].

Production costs in Canada tend to be lower overall, especially in the western provinces. The main reason is the cost of feed grains. In fact the advantage conferred by feed grain prices in the Prairie region is quite substantial. Another advantage is the Canadian interest rates.

Dr. MacMillan (Department of Agribusiness and Agricultural Economics) proposed a research project dealing with the economic impacts of hog production in Manitoba. The major topics in his project consist of:

1. Municipal decision making with respect to approving hog operation
2. Pembina Valley hog operation survey results
3. The Manitoba/ROC Input Output Model
4. Economic and environmental impacts of hog production

[2, 3, 4] www.gov.mb.ca/agriculture/statistics

5. Additional regional research needed on modeling, environmental, technical change, and extension for future development of hog production.

A more detailed proposal will be included in the Appendix. The Pembina Valley hog operation survey results will be the main focus of this paper.

The Pembina Valley is located in Manitoba and consists of eight municipalities:

1. Dufferin
2. MacDonald
3. Montcalm
4. Morris
5. Rhineland
6. Roland
7. Stanley
8. Thompson

(A map is found in the Appendix A)

Pembina Valley stretches over about 5000 square kilometres of prime agricultural land, which runs over the southern part of Manitoba.

In these 8 municipalities there are 217 hog producers. The producers consist of 151 small producers, 58 medium producers and 8 large producers. A small producer has 299 or less animal units, a medium producer has 300 to 1000 animal units and large producers consist of more than 1000 animal units. Due to cost and time constraints a sample of small producers was taken. A census of the medium producers and large producers were taken. Each producer received a

questionnaire (Appendix B). This questionnaire consists of 8 sections (all dealing with 2002 figures):

1. Type of hog production
2. Total Operating Cost
3. Source of Water
4. Employment for hog operation
5. Employees that reside in local Municipality
6. Breakdown of Household expenditures
7. Construction Expenditures
8. Brief Description of operation: i.e. Number of barns; manure system; expansion plans

Type of hog operation contains information such as type of business operation i.e. Do they own the hog operation solely or do they have a partner or are they part of a corporation? It also contains information on the number of hogs and how many hogs sold in 2002. Most importantly it indicates the type of hog operation. In the questionnaire there are 6 different types of hog operation:

1. Farrow to finish
2. Farrow to wean
3. Farrow to iso-wean
4. Feeder
5. Breeding stock
6. Custom Finishing

Farrow to finish operations include breeding pigs and are marketed at about 220 pounds. Farrow to wean operations include breeding pigs and are marketed at about 50 pounds. Farrow to iso-wean includes young pigs and are marketed at about 25 pounds. Feeders are fed from 25 to 50 pounds and marketed at about 220 pounds. Breeding stock operators raise sows (female hog) and boars (male hog) for breeding purposes. They sell sows and boars to other operators. The other operators get small piglets from the sows and boars they buy from breeding stock operators. Breeding stock operators do not sell their product to market directly. Custom finishing operators feed hogs from 25 to 50 pounds and marketed then at about 220 pounds. The particular type of operation will be important in the estimation procedure. This will be discussed in a latter chapter.

1.3 Further research

Imputation is a statistical technique for analyzing incomplete data sets, that is, data sets for which some entries are missing. Multiple imputation for this data is an interesting topic that can be further studied. In this paper single imputation will be used with the results from the survey to estimate economic impact for the Pembina Valley Region. Single imputation was used to substitute a value for each producer that did not respond to the survey. Proportion, ratio and regression estimates were used to obtain the missing values. This approach treats missing values as if they are known in the complete data analysis.

Chapter 2

Sampling

2.1 Sampling

The most informative sample is the whole population (i.e. a census). Due to time and cost constraints and accuracy it is usually not possible to take a census. The next best sample would be the target population. The target population is the complete collection of observations we want to study (Lohr, 1999). To maximize the amount of information for a certain cost we came up with a desired target population for the Pembina Valley Region. The population of hog producers were broken down into 8 different municipalities (strata):

Dufferin, MacDonald, Montcalm, Morris, Rhineland, Roland, Stanley and Thompson.

The animal units (AU) for each producer was available from the municipal office prior to drawing the sample, hence the population of hog producers can be further broken down into 3 subgroups:

Small: hog producer < 300 animal units;

Medium: 300 animal units $\leq$ hog producer ≤ 1000 animal units;

Large: hog producer > 1000 animal units;

The number of animal units in each of the subgroups was determined from a previous study for the municipality of MacDonald.

In previous studies (Pellow, 2000) and a pilot study of the municipality of Hanover it was suggested that the type of enterprise was also important. However this information was only available after approaching the producers.

Previous calculation Johnston (2000) showed how to determine sample size. His computations can be found in Appendix C. He also had the census list for a number of municipalities that include the number of animal units for each producer. He illustrated a procedure for determining sample size using incomplete information for the municipality of Stanley. To determine sample size the variance of labour of Stanley is needed. From a previous study (Pellow, 2000) information for the municipality of MacDonald was available. Johnston took the information for MacDonald to compute the sample size for Stanley. The results will be accurate if MacDonald results reflect Stanley's conditions. If the farm practice is quite different in Stanley than MacDonald, or other municipalities, them the sample size determined might not be sufficient to estimate the totals as precisely as desired.

There are a large amount of small producers in the Pembina Valley region. Since the variability of this group is relatively small (see Appendix C) a sample of 2 was suggested. After discussing it with the project team, a sample of 4 could be taken since it fit into the budget. It was also suggested to take a census of the medium and large hog producers. Three reasons were that there were a small number of them, we had the funding to do so, and the variability in these groups was high (see Appendix C). The variances for the 3 subgroups are quite different.

The impact of the overall variance will be mainly from the large group. A large variance will imply that a large sample should be taken. A census was suggested. For the small subgroup, a stratified random sample from the Pembina Valley region was obtained by selecting a simple random sample of four producers from each municipality. In total our desired sample for the small subgroup is 32 (four from each of the eight municipalities). There are a total of 58 medium hog producers and there are a total of 8 large hog producers in the Pembina Valley region.

2.2 Stratified Sampling

In stratified random sampling we divide the population into groups that do not overlap. A sampling unit (a unit we actually sample) only belongs to one stratum. We take a simple random sample from the stratum, and then use this information to estimate i.e. the population totals. We use stratified sampling for a few reasons:

1. Stratified sampling leads to efficient estimates. If we take a simple random sample of 100 hog producer from the Pembina Valley region, there is a chance that we would not obtain a sampling unit from a municipality with a small number of producers. Then that sample would not be a good representation of the population. By stratifying each group, we can indeed get information on all the groups.

2. Stratified sampling can give a precise estimate of the whole population. Hog producers in different parts of the Pembina Valley region will have different

characteristics, so it is recommended to stratify into different groups. Within each group there tends to be a lower variability in the whole population.

The following table illustrates the total number of hogs producers identified, the number of hog producers in each of the three subgroups and the number of small hog producers surveyed.

Table 2.1: Pembina Valley Hog Operations, Total Identified, Number of Small, Medium, Large Identified, and the Number of Small Hog Producers surveyed.

Stratum	Total No. of Hog Producers Identified	No. of Small Hog Producers Identified	No. of Small Hog Producers Surveyed	No. of Medium Hog Producers Identified	No. of Large Hog Producers Identified
1	17	16	4	*	*
2	18	10	4	*	*
3	9	6	4	*	*
4	50	26	4	20	4
5	82	64	4	18	*
6	8	5	4	*	*
7	26	20	4	5	*
8	7	4	4	*	*
Total	217	151	32	58	8

* Due to confidentiality reasons the data are too small to report or missing information

2.3 Notation for Stratified Sampling

The population of size N are divided into h strata, with size $N_1, N_2, \ldots,$ and N_h respectively; the strata are the rural municipalities of the Pembina Valley region Dufferin, MacDonald, ..., and Thompson respectively. The procedure is general and works with various variables. An example is shown in chapter 4.

The values N_1, $N_2, \ldots,$ and N_h are known therefore we know

$N_1 + N_2 + \ldots + N_h = N$,where N is the total number of producers in the Pembina Valley region.

Let N_{hj} = the number of producers of size j and in stratum h

j = small (s), medium (m) or large producer (l)

$N_{1s} + N_{1m} + N_{1l} = N_1$

The number of producers in the Dufferin municipality is equal to the sum of the number of producers in the small group, medium group and the large group.

x_{hjk} = number of animal units of the *kth* observation (i.e. hog operation) in size j and in stratum h (municipality)

y_{hjk} = value of the *kth* observation in size j and in stratum h (municipality)

$\tau_{hj} = \sum_{k=1}^{N_{hj}} y_{hjk}$ = population total value in size j and in stratum h (municipality)

$\tau_h = \sum_{j=1}^{J} \tau_{hj} = \sum_{j=1}^{J} \sum_{k=1}^{N_{hj}} y_{hjk}$ = population total value in stratum h (municipality)

$\tau = \sum_{h=1}^{H} \tau_h = \sum_{h=1}^{H} \sum_{j=1}^{J} \sum_{k=1}^{N_{hj}} y_{hjk}$ = population total value

$$\bar{y}_{hjU} = \frac{\tau_{hj}}{N_{hj}} = \frac{\sum_{k=1}^{N_{hj}} y_{hjk}}{N_{hj}} = \text{population mean in size } j \text{ in stratum } h \text{ (municipality)}$$

$$\bar{y}_{hU} = \frac{\tau_h}{N_h} = \frac{\sum_{j=1}^{J}\sum_{k=1}^{N_{hj}} y_{hjk}}{N_h} = \text{population mean in stratum } h \text{ (municipality)}$$

$$\bar{y}_U = \frac{\tau}{N} = \frac{\sum_{h=1}^{H}\sum_{j=1}^{J}\sum_{k=1}^{N_{hj}} y_{hjk}}{N} = \text{overall population mean}$$

$$S_{hj}^2 = \sum_{k=1}^{N_{hj}} \frac{\left(y_{hjk} - \bar{y}_{hjU}\right)^2}{N_h - 1} = \text{population variance in size } j \text{ and stratum } h$$

(municipality)

Using simple random estimates within each stratum,

the mean is

$$\bar{y}_{hj} = \frac{\sum_k y_{hjk}}{n_{hj}}$$

the total is

$$\hat{\tau}_{hj} = \frac{N_{hj}}{n_{hj}} \sum_k y_{hjk} = N_{nj}\, \bar{y}_{hj}$$

the sample variance is

$$s_{hj}^2 = \sum_k \frac{\left(y_{hjk} - \bar{y}_{hj}\right)^2}{n_{hj} - 1}.$$

In the chapters 4 and 5 there will be illustrations in finding population totals and variances.

Chapter 3

Nonresponse and Imputation

3.1 Nonresponse

An ideal survey is to have no nonresponse. A nonresponse is failing to obtain all the responses from the target sample. One way to handle this problem is to predict the missing information. One advantage in our survey is that we do have some information on the producers who did not respond. We have information on their animal units and the type of operation they have (i.e. Farrow to wean etc.). Of course, predicting the missing data is not as good as observing them. The prediction process introduced another source of imprecision.

3.2 Unit and Item Nonresponse

There are two types of nonresponse: unit nonresponse, in which the producers did not respond at all, and item nonresponse, in which some information in the questionnaire was not completely filled out (i.e. A particular question was not answered). For producers who did not fill out the questionnaire at all, it might be that they were on a trip and could not respond, or they refuse because they do not want their information released. Our survey had questions on revenue, cost, etc., and it would take some time to go through their files (tax return, financial statements, and receipts) to answer these questions. They might not have the time to do so. The number of animal units from each of the producers was known.

This information was available from the municipality office. The type of operation was not known at the start of the project but later assessed by the observer. Item nonresponse occurred because of refusal of answering a particular question. A producer might object to answering personal questions i.e. income and/or revenue.

3.3 Dealing with Nonrespondents

We worked with nonrespondents with great caution. First, we tried to prevent it by designing a questionnaire that will keep the number of nonrespondents low. Second, we made estimates by ignoring the nonresponse but realize this method has its weakness. This would create potential nonresponse bias. Our estimate may be inaccurate if we ignore the nonresponse. Third, we used models to predict the values of the nonresponse. Both ratio estimates and regression analysis were used to adjust for unit nonresponse. Imputation was used to adjust for item nonresponse. Imputation is when you assign a value to the missing data. Regression analysis was used to predict nonresponse values and these imputed values were used in the estimation.

In our study, there was an overall 98 questionnaires sent out and 56 responded. We took a sample of the small producers and a census of medium and large producers. Twenty one questionnaires out of the 32 sent out to the small producers were returned. For the medium, we received 28 questionnaires out of the 58 sent out. We sent out 8 questionnaires for the large producers and received

7 responses. As you can see, there is a problem for medium producers. More than half of the total questionnaires were not returned.

There are many reasons why there was a high nonresponse rate. These reasons can be related to the characteristics of the questionnaire, respondents and the interviewer. The questionnaire might have sensitive questions to the producer, such as questions on income and other financial matters. The questionnaire might have been too long and therefore may have been incomplete or just put aside to be finished at a later time and just forgotten.

Some large producers might not have released information for competitive reasons. Another factor to consider is the respondents' characteristics. They may not have the time and they may not have the motivation to answer the questionnaire. Producers answering the questionnaire are doing us a favor. We used a short questionnaire. Most of the information in the questionnaire is needed for the input-ouput model. Some additional questions were added to predict some data that is required for this model. The data collection method of mailing the questionnaires is considered a good idea. This type of method has been proven to achieve a low nonresponse rate.

In our research project we had only one interviewer, Slyvio Sabourin. He was responsible for mailing and collection of the questionnaire. If the questionnaire were not received after a certain period of time, Slyvio would follow up with the producer. For the small producers, if they did not respond after the follow up, a substitution was made. A random sample of small producers was taken and additional surveys were sent to those producers.

3.4 Imputation

Nonresponse can be handled by replacing the missing values by a number based on the information on hand. In many statistical procedures, imputation is commonly used to assign values to this missing information. For these missing values we predict values from producers whose information is complete. This will reduce nonresponse bias.

3.4.1 Cell Mean Imputation

The mean of the respondents in each municipality for each subgroup is calculated. For example the mean of the number of employees for all the small producers in the Dufferin municipality was calculated. This value was then imputed for all the nonrespondents in the sample prorated as to operation size of animal units. A major assumption is that nonrespondents would answer in a similar fashion as producers with the same operation size and municipality that did respond.

3.4.2 Regression Imputation

Using regression analysis we can predict missing values. The missing values were replaced by a predicted value from a regression model. Equations were developed for the producers with complete information using animal units and a coded variable: farm type. An example is shown in Chapter 4 section 4.4.

3.4.3 Substitution

As mentioned in the previous chapter, the interviewer was allowed to randomly choose a substitute in the same municipality for the small producers. For example, if a questionnaire was not returned by a certain time, the interviewer was to randomly choose three different producers to receive the questionnaire, to make sure we have four small producers in each municipality. This helps in reducing some nonresponse bias since other small producers may have the same characteristics as the small producers selected. A disadvantage may be that the sample no longer has known probabilities of selection. The probabilities for each municipality are now different.

3.5 Advantage and Disadvantage

3.5.1 Advantage

Primarily, imputation creates a complete data set. This will help reduce or eliminate nonresponse bias. By having a complete data set, it is easier and simpler to work with. When all the missing values have been filled in, standard methods of analysis can be done. The variance of the estimates can be obtained.

3.5.2 Disadvantage

Imputation treats missing values as known data values. This will cause the sample variance to be too small. The appropriate variance of the estimate will now be sampling variance plus the prediction variance for the values that were imputated. This considerably increases the estimate variance, since we are using

estimated values rather than observed values. Uncertainty is introduced therefore the variance estimation is increased by the prediction variance.

The Canadian census is an example where imputation was used. Statistics Canada has been using the method of imputation for quite some time. The census, for example, has incomplete data. Therefore, Statistics Canada will find a similar response with similar characteristics and fill in the missing value with the data from the actual respondents.

Chapter 4

Results

There were three estimation approaches used in this project. They are proportion, ratio, and regression estimation. Observations were taken on a sample of small hog producers and a census of medium and large hog producers, by y_{hjk}, the *kth* observation in size *j* and for producer *h*. Also x_{hjk} represents the number of animal units. We obtain the x_{hjk}'s for each producer in the municipalities. We assume that the number of employees is related to the number of animal units. The hog study (MacMillan et al.) included a number of other variables (i.e. income).

There was some nonresponse: i.e. producers that did not respond to the survey. These producers were adjusted using proportion, ratio and regression estimation. In the following illustrations labour is the variable of interest.

4.1 Basic notation

The population total for the variable of interest for each stratum, which is denoted by τ_H, is the sum of all observations in the stratum. The total labour force in the region, which is denoted by τ_H, is the objective of the study. The population total is obtained by finding the small, medium, and large producer totals, which are

denoted by $\tau_{hs}, \tau_{hm}, and\, \tau_{hl}$ respectively. The total is calculated by adding all the subgroup totals.

$\tau_{H} = \tau_{hs} + \tau_{hm} + \tau_{hl}$

4.2 Proportion Estimation for the Small and Medium Producers Total

N_{hs} is the number of small producers in the municipality. The population total of the small producers is τ_{hs} and its estimate is $\hat{\tau}_{hs}$. Hence

$\hat{\tau}_{hs} = N_{hs}\,\overline{y}_{hs}$ where

$\overline{y}_{hs}$ is the mean labour force of small producer

Due to a high nonresponse rate for medium producers (52%), there was a decision to estimate the total of medium producers τ_{hm} using an adjustment i.e.

$$\hat{\tau}_{hm} = \frac{N_{hm}}{n_{hm}} \tau_{hm(received)}$$

In this case, $\tau_{hm(received)}$ is the total number of employees from the medium producers who actually responded. The total number of medium producers in the population is denoted by N_{hm}. The total number of medium producers that actually responded is denoted by n_{hm}. The following table illustrated an example of proportion estimation.

4.2.1 Illustration of proportion estimation

Table 4.1: An Example of Proportion Estimation

Survey	AU	Employees
1	84	0.42
2	84	0.41
3	150	2.00
4	312.50	6.15
5	375	4.57
6	385	1.20
7	429	5.00
8	1285.34	2.14

Notes:

1. Data received from surveys
2. Survey ordered by number of animal units
3. Full time equivalent employees include paid employee, unpaid family and management.
4. Animal units is denoted by AU
5. The first three observations (observations 1-3) are small producers that have less than 300 AU
6. The next four observations (observations 4-7) are medium producers that have 300 AU or greater but equal to 1000AU or less
7. The last observation (observation 8) is a large producer that has more than 1000 AU

Thus using proportion estimation we estimate the total labour force as

$$\hat{y}_h = N_{hs}\,\bar{y}_{hs} + \frac{N_{hm}}{n_{hm}}\tau_{hm} + \tau_{hl} \qquad \textbf{(4.1)}$$

where

N_{hs} is the total number of small producers within municipality h;

$$\bar{y}_{hs} = \frac{\sum_{h=1}^{H} y_h}{n_{hs}}$$

y_h is the number of employees within municipality h

n_{hs} is the total number of small producers that replied within municipality h

$\bar{y}_{hs}$ is the mean employees of the small producers that have replied within municipality h

τ_{hm} is the total number of employees in the medium sized category on farms that have replied within municipality h

N_{hm} is the total number of medium producers within municipality h

n_{hm} is the number of medium producers that replied within municipality h

τ_{hl} is the total number of employees for the large sized producers within municipality h

In the example data of table 4.1

N_{hs}=20 where n_{hs}=3 since 3 small producers replied

Using the formula (4.1)

$$\bar{y}_{hs} = \frac{\sum_{h=1}^{H} y_h}{n_{hs}} = \frac{y_{h1} + y_{h2} + y_{h3}}{3} = \frac{0.42 + 0.41 + 2.00}{3} = \frac{2.83}{3} = 0.94$$

$N_{hm} = 5$

$n_{hm} = 4$

$$\tau_{hm} = \sum_{k=1}^{K} y_{hmk} = y_{hm1} + y_{hm2} + y_{hm3} + y_{hm4}$$

$$= 6.15 + 4.57 + 1.20 + 5.00$$

$$= 16.92$$

$$\tau_{hl} = \sum y_{hl} = y_{hl} = 2.14$$

Now the estimate total number of employees, $\hat{y}$ is:

$$\hat{y} = N_{hs}\bar{y}_{hs} + \frac{N_{hm}}{n_{hm}}T_{hm} + T_{hl}$$

$$= 20(0.94) + \frac{5}{4}(16.92) + 2.14$$

$$= 18.8 + 21.15 + 2.14$$

$$= 42.09$$

We estimate 42 employees in this municipality.

The medium producers were adjusted since there was some nonresponse.

4.3 Ratio Estimation

Another method used to estimate the total number of employees is Ratio estimation. We adjust the nonresponse by multiplying the ratio $\frac{\overline{y}}{\overline{x}}$ (using the data only from the respondents) by the population total τ_t. Since there is a relationship between y and x, ratio estimation can be used (previous studies it shows this a linear relationship by y and x). This method requires a measurement of two variables y_{hj} and x_{hj}. The population of the small producers has total employees, τ_{hs} in stratum h. Hence using ratio estimation we can estimate τ_{hs} by:

$$\hat{\tau}_{hs} = \frac{y_{hs}}{x_{hs}} X_{hs}$$

where

x_{hs} is the total number of animal units of small producers in the sample (received) for the municipality

X_{hs} is the total number of animal units in the municipality from small producers

Due to a large number of nonrespondents of medium producers, we have adjusted the total for medium producers, namely

$$\hat{\tau}_{hm} = \frac{x_{\text{missing}}}{x_{received}} \tau_m + \tau_m$$

$x_{missing}$ is the total number of animal units of the nonrespondents of the medium producers and

$x_{received}$ is the total number of animal units of the sample respondents of the medium producer.

The following table is an illustration of ratio estimation.

4.3.1 Illustration of Ratio Estimation

Table 4.2: An Example of Ratio Estimation

Producer	AU	Employees
1	84	0.42
2	84	0.41
3	150	2.00
4	312.5	6.15
5	375	4.57
6	385	1.20
7	429	5.00
8	1285.34	2.14
9	2.5	
10	30	
11	30	
12	35	
13	40	
14	43	
15	48	
16	57	
17	94	
18	100	
19	115	
20	129	
21	143	
22	172	
23	200	
24	252	
25	286	

Notes:

1. One full time employee is equivalent to 1 and 1 part time employee is equivalent to 0.5 employees etc.
2. Data ordered by size of animal units within respondents and nonrespondents
3. The first eight pieces of data represents the producers that actually handed in a survey, while observations 9 to 25 represent the producers that refused, or not included in the sample (nonrespondents)
4. Animal units is denoted by AU
5. The first three observations (observations 1-3) are small producers that have less than 300 AU
8. The next four observations (observations 4-7) are medium producers that have 300 AU or greater but equal to 1000AU or less
6. The last observation (observation 8) is a large producer that have more than 1000 AU

The predicted number of employees may be calculated as

$$\hat{y}_h = \frac{y_{hs}}{x_{hs}} X_{hs} + \frac{x_{h\ missing}}{x_{h\ received}} \tau_{hm} + \tau_{hm} + \tau_{hl} \qquad \textbf{(4.2)}$$

$\hat{y}_h$ is the predicted number of employees in stratum h

x_{hs} is the total number of animal units of small producers in the survey

X_{hs} is the total number of animal units of small producers in the municipality

τ_{hm} total number of employees in the medium sized category that have replied

$x_{h\ missing}$ is the total number of animal units of the missing surveys (medium producers)

$x_{h\ received}$ is the total number of animal units of the received surveys (medium producers)

τ_{hl} is the total number of employees of the large sized producers. Now

$$y_{hs} = 0.42 + 0.41 + 2.00 = 2.83$$

$$x_{hs} = 318$$

$$X_{hs} = 2094.5$$

$$\tau_{hm} = 16.92$$

$$x_{h\ missing} = 0$$

$$x_{h\ received} = 1502$$

$$\tau_{hl} = 2.14$$

Using formula (4.2)

$$\hat{y}_h = \frac{y_{hs}}{x_{hs}} X_{hs} + \frac{x_{h\ missing}}{x_{h\ received}} \tau_{hm} + \tau_{hm} + \tau_{hl}$$

$$\hat{y}_h = \frac{2.83}{318} 2094.5 + \frac{0}{1502} 16.92 + 17.48 + 2.14$$
$$= 18.63 + 0 + 16.92 + 2.14$$
$$= 37.69$$

4.4 Regression estimation

Regression estimation is the third method used in determining properties for variables for this project. Regression estimation gives us a little more information about the variable y from the information provided by the variable x. If there is evidence of a linear relationship between the observed y's and x's regression estimation procedure will be effective. This procedure is an improvement on the ratio estimate particularly when the line does not pass through the origin.

4.4.1 Regression estimation example

The model

$$y_k = \beta_0 + \beta_1 X_{k1} + \beta_2 X_{k2} + \beta_3 X_{k3} + \varepsilon_k$$

where

y_k is the number of employees

X_{k1} is the number of animal units

$$X_{k2} = \begin{cases} 1 & if \quad type^* \quad is \quad other \\ 0 & otherwise \end{cases}$$

$$X_{k3} = \begin{cases} 1 & if \quad type^* \quad is \quad large \\ 0 & otherwise \end{cases}$$

*There are 3 categories of hog production: other, large and farrow to iso-wean. The 'other' group contains all producers that were not large producer (producers