Study of the estimation precision in the Chinese MIS surveys

Bengt Rosén

SCB

R&D Report
Statistics Sweden
Research - Methods - Development
1991:5

Från trycket	Mars 1991
Producent	Statistiska centralbyrån, utvecklingsavdelningen
Ansvarig utgivare	Åke Lönnqvist
Förfrågningar	Bengt Rosén, 08-783 44 90

ISSN 0283-8680
Printed in Sweden

INLEDNING

TILL

**R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.
Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.**

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1991:5. Study of the estimation precision in the Chinese MIS surveys / Bengt Rosén.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1991-X101OP9105

Study of the estimation precision in the Chinese MIS surveys

by

Bengt Rosén

The contents of this paper has earlier been presented as part of the Mission Report: KINSTAT 1991:1, International Consulting Office, Statistics Sweden.

Abstract

Since some years the Chinese State Family Planning Commission has worked to establish a sample-based information system for monitoring the impact of the national family planning policy, a system referred to as MIS. Statistics Sweden cooperates in the project. Sofar the MIS surveys have been viewed as pilot studies intended to yield information on and experience of various empirical matters of relevance for the final design of the system. MIS was started at the prefecture level, and has gradually been extended so that pilot study data now are available for entire provinces.

The paper reports on a study of the sampling errors for statistics from the MIS, which are evaluated on the basis of empirical MIS survey data. In particular the study is concerned with the estimation precision of statistics on the prefecture level as well as on the province level.

Statistics Sweden
International Consulting Office (ICO)

February 1991

Study of the estimation precision in the Chinese MIS surveys

Bengt Rosén

1. Introduction

The Chinese State Family Planning Commission is developing a "computerized information system for monitoring the impact of the Chinese family planning efforts", and Statistics Sweden/ICO cooperates in the project. This system, referred to as MIS, shall provide information on the state of and the evolution over time of various statistical parameters of relevance for the Chinese family planning program. The basic types of parameters of interest in the MIS context are as follows;

- **Total numbers** of individuals in various types of groups (e.g. the number of women who use IUD contraceptives, the number of children born during a year).

- **Proportions** of different kinds of attributes (e.g. kind of contraceptive method used) in different population groups (notably different types of groups of women).

- **Rates/frequencies** of different kinds of events, notably births, (e.g. crude birth rate, general fertility rate) for different population groups.

- Indices of **change over time** (in e.g. the crude birth rate).

As indicated above, the MIS statistics shall cover a variety of types of groups. Of particular relevance are groupings according to geographical/administrative criteria, and regions of special interest are **prefectures, provinces** and the **whole country**.

In its broad lines the MIS system runs as follows. **Samples** of individuals are observed with respect to variables of the types which are indicated above, and statistical parameters are estimated from the sample data. As always for sample based statistics, the statistics from MIS are afflicted with **uncertainty** due to the restriction to samples, or in other words the statistics are afflicted with **sampling errors**. The desires as regards **estimation precision** are (to my knowledge) stated as follows;

- Estimates of (the majority of) the proportions of interest should have good precision for prefectures.

- Estimates of the rates of interest should have good precision on the level above prefectures, i.e. for provinces.

The achievement of desired estimation precisions depends in particular on the sampling design which is employed for the MIS surveys, notably their sample sizes. The final design for the MIS surveys is still to be decided upon. Accordingly, the first phase in the implementation of the MIS system has been viewed as a **pilot survey phase**, with the main aim to gain information on various empirical aspects of the surveys, i.a. empirical characteristics which enable evaluation of the precision of the statistics from the MIS surveys. "First guesses" of the sampling errors can be given "at the desk" (without empirical data), and have been presented to give guidelines for the planning during the pilot survey phase. Such "guestimates" of the order of magnitude of the sampling errors can be found in i.a. the reports KINASTAT 1988:2 (by Thomas Polfeldt) and KINASTAT 1990:1 (by Ola Nygren). However, "desk guestimates" of estimation precisions can only be tentative, as they inevitably have to be based on **assumptions** about the variability pattern in the population.

By now, data from the pilot surveys are available and in particular there are MIS data from all prefectures in the Liaoning province. These data allow empirical evaluation of the sampling errors for prefecture estimates as well as for province estimates. The main objectives of the present report are as follows.

To **report on empirical values** for the margins of sampling error, derived from the MIS data from the Liaoning province. Notably, empirical values for the estimation precision for prefecture statistics as well as for province statistics are presented. It is also of interest to compare the empirically evaluated margins of sampling errors with previous "guestimates", and material for such comparisons is presented.

The rest of the report is organized as follows. In Section 2 we present some facts on the design of the MIS surveys. In Section 3 we describe the point estimation procedures which have been considered in the present study. This description is elaborated on in Section 5, where we also specify the procedures which were used in the evaluation of the margins of sampling errors. The numerical findings are presented and discussed in Section 4.

2. Some basics facts on the MIS surveys

The subsequent description of the sampling and estimation procedures in the MIS surveys relates to the design which was used in the hitherto pilot surveys. A basic feature of the sampling design is that individuals are sampled in **clusters** by employing so called **area sampling**. More information is provided below.

Formation of sampling areas: China is divided into some 30 **provinces**, and each province is divided into a number of (around 10) **prefectures**. The prefectures are hierarchically divided into smaller areas, which are disjoint on their respective levels in the partitioning chain. At the end of this subdivision process come so called **village groups**, henceforth abbreviated VG. Orders of magnitude for the populations in the mentioned types of areas are as follows; A province contains on the average 36 million people, a prefecture 3.3 millions and a VG around 200 individuals.

VGs are used as area sampling units, and sampling frames for VGs are provided by listings which were made in conjunction with the "2 in 1000" Fertility survey, 1988.

Sampling procedure: Within a prefecture a random sample of VGs is selected by systematic sampling. Before the selection of the sample, the VGs are grouped into five different "**sectors**", also called strata, depending on their degree of urbanization ("urban", "rural", "farm", "township", "outskirts"). Within the sectors the VGs are listed in "geographical order". (We use the term "sector" instead of "stratum" to avoid confusion with the fact that the prefectures function as "sampling strata" in the provinces.) The number of VGs to be selected in a prefecture, technically the "sampling interval" in the systematic sampling, was chosen so that the total number of individuals within sampled VGs should be of prescribed size, at least roughly. In the pilot surveys the aimed at population size was 10 000 individuals. In the selected VGs, information on all individuals is collected, i.e. no sub-sampling within VGs is employed.

Data collection: When establishing the MIS system within a selected VG, an initial round is carried out in which information is collected via questionnaires which are filled out by interviewers who visit the households in the VG. This information concerns basically various vital statistics data, and information on present as well as past conditions and events is collected. The initial information is then updated every months, by collection and registration of information on changes since "last month". Certain general information relating to the VGs, as e.g. their degree of urbanization, is collected from VG agencies, but all demographic information is collected directly from the households.

3. On point estimation procedures

Below we specify the point estimation procedures which are considered in the present study. Unfortunately it has been difficult to get unequivocal information on the point estimation procedures in the actual MIS system, and maybe the following specification means a slight simplification of matters. However, if so we judge that the effects of the (possible) simplifications are negligible as regards the main issue; Empirical values for the margins of sampling errors. The point estimation procedures used in the study are based on the following model for the probabilistic structure of the MIS samples.

The sample of individuals from a prefecture is a cluster sample, with the VGs as clusters, where the clusters are selected with equal probabilities. (1)

The point estimation procedures are specified more formally in Section 5 where also the estimation procedures for sampling errors are specified and discussed.

3.1 Estimates for prefectures

As a consequence of (1) the sample of individuals from a prefecture is **self-weighting**, i.e. all individuals in the prefecture have the same probability of being included in the sample. For a sample with this property (provided that the non-response rate is low, which is assumed), means, proportions and rates in the prefecture are estimated the "straightforward way", i.e. by the corresponding means, proportions and rates in the sample. The same applies for indices of change, as defined below.

$$\textbf{\textit{Index of change}}\ \textit{between year x and y} = \frac{\textit{proportion/rate for year x}}{\textit{proportion/rate for year y}}. \tag{2}$$

Total numbers (e.g. of eligible women, births, etc) in a prefecture are estimated by grossing up the corresponding sample number by the factor;

$$\textbf{\textit{Grossing up factor}} = \frac{\textit{Total number of VGs in the prefecture}}{\textit{Sampled number of VGs in the prefecture}}. \tag{3}$$

Aspects where the above specifications of the point estimation procedures may deviate from the actual MIS procedures are the following. Somewhat different systematic sampling intervals may be used in the different sectors, which would disturb the self-weighting property. The grossing up may be made on basis of current population figures, and not on basis of the sampling rates for VGs. However as already stated, we believe that the possible discrepancies have only marginal effects on the estimates, on the point estimates as well as on the estimates of sampling errors.

3.2 Point estimates for provinces

The sample of individuals from a province is not self-weighting, though. The prefectures function as sampling strata and the sampling rates vary considerably between prefectures (see Table 1). Therefore, statistical characteristics for provinces were estimated by first estimating relevant total numbers for the province and then forming estimates of means, proportions, rates, etc as ratios of the appropriate total number estimates. Province totals were estimated by employing the following principle;

Province totals are estimated as the sum of the estimates of the corresponding prefecture totals. (4)

4. Numerical findings on the estimation precision in the MIS surveys

4.1 The study data

The main task in this section is to present and discuss various numerical results on sampling errors, which are derived from the MIS data from the Liaoning province. Liaoning contains the following prefectures; Shengyang, Dalian, Anshan, Fushun, Benxi, Dandong, Jinzhou, Yingkou, Fuxin, Liaoyang, Tieling, Chaoyang, Panjin and Jinxi. Unfortunately, though, data for three of the prefectures were damaged in the transportation from Beijng to Stockholm, and therefore the prefectures Anshan, Panjin and Jinxi have been excluded. Accordingly, the province "Liaoning" which is reported on in the following, is a somewhat deduced version of the real province Liaoning to the effect that it does not comprise the three excluded prefectures. However, for the present illustration purpose, omission of three prefectures should not be very harmful. "Liaoning" still contains 11 prefectures, which in fact is a typical number of prefectures in a province. Hence, the reported results should give a good picture of the orders of magnitude of sampling errors for prefecture characteristics as well as for province characteristics. In fact we believe that the estimates for "Liaoning" lie quite close to those of the real Liaoning, regarding point estimates as well as estimates of the sampling errors.

We start with some general information on the samples from the "Liaoning" prefectures as well on the prefectures themselves.

Table 1. Some characteristics for the samples and the populations.

Region	Number of sampled VGs.	1989 population in the sampled VGs.	Total number of VGs in the region.	Total population 1988 (or 1987).
Pref Shengyang	90	9 957	49 175	5 115 236
Pref Dalian	59	9 740	30 195	4 685 727
Pref Fushun	75	10 073	15 675	2 516 835
Pref Benxi	61	9 568	9 113	1 402 240
Pref Dandong	50	9 511	13 892	2 600 790
Pref Jinzhou	50	9 880	23 290	4 467 909
Pref Yingkou	47	9 337	9 700	2 743 716
Pref Fuxin	51	9 222	8 811	1 694 294
Pref Liaoyang	68	9 721	9 132	1 590 170
Pref Tieling	48	9 795	18 361	3 379 144
Pref Chaoyang	40	10 343	14 085	3 465 488
Prov "Liaoning"	639	107 147	201 429	33 661 549

In the study we used MIS data relating to the years 1988 and 1989, and the data file for the study consisted of aggregated values over VGs, which were compiled from the basic MIS files. In most of the prefectures MIS was started up during 1989 and in such cases the 1988 data are retrospective, and their quality may be questioned on that ground. As regards the 1989 data, some of them might be slightly "defect" to the effect that all births during December are not included. We have made evaluations of the sampling errors for 1988 as well as 1989, and both years are reported on in Table 12, but we confine the detailed reporting to 1989 figures. Some of the (point) estimates for birth rates rates and changes of birth rates are admittedly "strange". This may depend on deficiencies in 1988 or in the 1989 data or on both. However, even if "strange" point estimates occur because of data deficiencies, we believe that their effect on the orders of magnitude of the sampling errors is much smaller and that the estimated sampling errors are of the right order of magnitude all the same.

4.2 Numerical findings on sampling errors

Numerical values for sampling errors are presented in the following Tables 2 - 11, which are organized with the following general structure (i) - (iv) (which goes from left to right in the tables).

(i) The **point estimate** for the particular statistical characteristic is presented.

(ii) The (absolute) **margin of sampling error** for the estimated value is reported. This value is computed as 2 × (the estimated standard deviation of the point estimator). Another way of expressing the same thing is to say that

$$\text{estimated value} \pm \text{margin of error} \tag{5}$$

yields an (approximately) **95% confidence interval** for the characteristic under consideration. The reported margins of error were computed as follows;

$$\textit{Margin of error} = 2 \cdot \sqrt{\hat{V}(\textit{estimator})} \tag{6}$$

where the estimate $\hat{V}$(estimator) of the estimator variance was computed as described in Section 5.

(iii) The value of the **relative margin of (sampling) error** is presented, computed as;

$$\textit{Relative margin of error} = \frac{\textit{Margin of error}}{\textit{Estimated value}}. \tag{7}$$

(4) "**Guestimates**" of the margins of sampling errors are presented, and their computations are based on simplifying assumptions, notably the following one;

The total sample of individuals from a prefecture, as well as that from a province, is regarded as a simple random sample of individuals. (8)

The computation formulas for the "guestimates" of the margins of sampling errors are presented in Section 5, where we also introduce another simplifying assumption. There are two reasons why the "guestimates" are of interest.

(i) As has been stated, the sample of individuals from a prefecture is a cluster sample, with VGs as clusters, and not a simple random sample. Under assumption (8), one disregards possible correlations between observations which emanate from the fact that individuals within a VG may behave more similarly than individuals in a totally random sample from the prefecture. However, there is reason to believe in fairly strong **intra cluster correlations** at least for some of the variables under consideration. One can surmise fairly strong correlations for variables which relate to contraceptive method, while it is reasonable to believe that events as births are weakly correlated within VGs. When such intra cluster correlations prevail, the assumption (8) leads to under-estimation of the sampling errors, i.e. to too "optimistic" error estimates. Comparison of actual sampling errors and those under assumption of simple random sampling (with the same sample size of individuals), often made via the notion of **design effect**, give information on the intra cluster correlations, and thereby on the efficiency of the sampling design.

(ii) As stated before, until recently no empirical MIS data have been available and in lack of such, needed judgements could only be made on the basis of assumptions, and the assumption (8) has been employed (i.a. in KINASTAT 1988:2 and KINASTAT 1990:1). Since "guestimates" have been used in earlier MIS considerations, it is of interest to see how good/bad they are.

Table 2. Number of eligible women 1989

Region	Estimated number	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	1 580 156	235 796	14.9%	3.1%
Pref Dalian	1 420 700	264 020	18.6%	3.2%
Pref Fushun	608 399	99 419	16.3%	3.1%
Pref Benxi	415 314	72 419	17.4%	3.2%
Pref Dandong	754 336	153 148	20.3%	3.2%
Pref Jinzhou	1 369 452	231 936	16.9%	3.1%
Pref Yingkou	526 689	135 323	25.7%	3.4%
Pref Fuxin	475 621	78 462	16.5%	3.2%
Pref Liaoyang	376 561	79 396	21.1%	3.2%
Pref Tieling	1 171 661	252 972	21.6%	3.0%
Pref Chaoyang	1 056 375	236 519	22.4%	3.1%
Prov "Liaoning"	9 755 265	606 953	6.2%	1.0%

Table 3. Number of eligible women who use contraceptives 1989.				
Region	Estimated number	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	1 202 056	180 989	15.1%	3.7%
Pref Dalian	1 085 996	221 004	20.4%	3.8%
Pref Fushun	384 769	61 635	16.0%	4.2%
Pref Benxi	275 930	44 691	16.2%	4.0%
Pref Dandong	554 013	117 671	21.2%	4.1%
Pref Jinzhou	904 118	163 526	18.1%	4.4%
Pref Yingkou	345 691	90 311	26.1%	4.3%
Pref Fuxin	297 155	50 902	17.1%	4.0%
Pref Liaoyang	263 216	53 547	20.3%	3.9%
Pref Tieling	805 971	176 317	21.9%	3.9%
Pref Chaoyang	771 631	154 893	21.6%	4.0%
Prov "Liaoning"	6 836 547	443 485	6.5%	1.2%

Table 4. Number of eligible women who used IUD 1989.				
Region	Estimated number	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	852 913	117 136	13.7%	4.6%
Pref Dalian	662 243	129 285	19.5%	5.2%
Pref Fushun	273 372	34 307	12.5%	5.2%
Pref Benxi	182 858	30 229	16.5%	5.3%
Pref Dandong	318 682	76 105	23.9%	5.5%
Pref Jinzhou	516 106	86 499	16.8%	5.7%
Pref Yingkou	199 572	49 686	24.9%	6.1%
Pref Fuxin	189 350	29 575	15.6%	5.7%
Pref Liaoyang	172 837	32 518	18.8%	5.2%
Pref Tieling	484 271	97 694	20.2%	5.2%
Pref Chaoyang	376 422	129 292	34.3%	5.8%
Prov "Liaoning"	4 228 626	276 518	6.5%	1.6%

Table 5. Number of children born 1989.				
Region	Estimated number	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	68 845	16 432	23.9%	17.7%
Pref Dalian	68 067	16 426	24.1%	17.2%
Pref Fushun	28 424	8 679	30.5%	17.0%
Pref Benxi	13 744	3 656	26.6%	20.8%
Pref Dandong	37 231	11 210	30.1%	17.2%
Pref Jinzhou	41 922	12 051	28.7%	21.0%
Pref Yingkou	29 926	8 887	29.7%	16.5%
Pref Fuxin	14 685	3 715	25.3%	21.6%
Pref Liaoyang	15 309	4 786	31.3%	18.6%
Pref Tieling	44 754	15 068	33.7%	18.4%
Pref Chaoyang	47 889	14 292	29.8%	17.0%
Prov "Liaoning"	410 796	38 032	9.3%	5.5%

Table 6. Contraceptive prevalence among eligible women 1989.				
Region	Estimated proportion	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	76.1%	3.5%	4.6%	2.1%
Pref Dalian	76.4%	5.1%	6.7%	2.1%
Pref Fushun	63.2%	3.0%	4.8%	2.8%
Pref Benxi	66.4%	2.8%	4.1%	2.7%
Pref Dandong	73.4%	4.2%	5.7%	2.3%
Pref Jinzhou	66.0%	3.6%	5.4%	2.6%
Pref Yingkou	65.6%	2.1%	3.1%	2.9%
Pref Fuxin	62.5%	2.7%	4.3%	3.0%
Pref Liaoyang	69.9%	3.3%	4.8%	2.5%
Pref Tieling	68.8%	2.8%	4.0%	2.4%
Pref Chaoyang	67.9%	2.9%	4.2%	2.5%
Prov "Liaoning"	70.1%	1.3%	1.8%	0.7%

Table 7. Proportion IUD users among eligible women 1989.				
Region	Estimated proportion	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	54.0%	4.9%	9.0%	3.4%
Pref Dalian	46.6%	4.4%	9.4%	4.1%
Pref Fushun	44.9%	4.7%	10.4%	4.1%
Pref Benxi	44.0%	4.4%	9.9%	4.3%
Pref Dandong	42.2%	5.7%	13.5%	4.5%
Pref Jinzhou	37.7%	3.8%	10.2%	4.7%
Pref Yingkou	37.9%	3.9%	10.3%	5.1%
Pref Fuxin	39.8%	3.4%	8.5%	4.7%
Pref Liaoyang	45.9%	3.9%	8.6%	4.1%
Pref Tieling	41.3%	3.0%	7.2%	4.3%
Pref Chaoyang	35.6%	6.0%	16.9%	4.9%
Prov "Liaoning"	43.3%	1.5%	3.4%	1.3%

Table 8. Crude Birth Rate (CBR) 1989.				
Region	Estimated rate	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	1.27%*	0.21%	16.4%	17.7%
Pref Dalian	1.37%*	0.23%	16.7%	17.2%
Pref Fushun	1.35%*	0.26%	19.6%	17.0%
Pref Benxi	0.96%*	0.19%	19.8%	20.8%
Pref Dandong	1.40%*	0.25%	17.4%	17.2%
Pref Jinzhou	0.91%*	0.21%	22.6%	21.0%
Pref Yingkou	1.55%*	0.34%	21.8%	16.5%
Pref Fuxin	0.92%*	0.17%	18.9%	21.6%
Pref Liaoyang	1.17%*	0.23%	19.4%	18.6%
Pref Tieling	1.19%*	0.27%	22.7%	18.4%
Pref Chaoyang	1.31%*	0.27%	20.6%	17.0%
Prov "Liaoning"	1.23%*	0.08%	6.4%	5.5%

*) Recall the comments on page 6 concerning possible data deficiencies.

Table 9. General Fertility Rate (GFR) 1989.				
Region	Estimated rate	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	4.4%*	0.7%	16.3%	17.4%
Pref Dalian	4.8%*	0.8%	17.2%	16.9%
Pref Fushun	4.7%*	0.9%	20.3%	16.7%
Pref Benxi	3.3%*	0.7%	19.7%	20.5%
Pref Dandong	4.9%*	0.8%	17.0%	16.8%
Pref Jinzhou	3.1%*	0.7%	23.4%	20.8%
Pref Yingkou	5.7%*	1.3%	23.1%	16.1%
Pref Fuxin	3.1%*	0.6%	19.6%	21.4%
Pref Liaoyang	4.1%*	0.8%	18.9%	18.3%
Pref Tieling	3.8%*	0.9%	24.7%	18.1%
Pref Chaoyang	4.5%*	0.9%	21.0%	16.8%
Prov "Liaoning"	4.2%*	0.3%	6.6%	5.4%

Table 10. Index of change in the proportion of IUD users between 1988 and 1989.				
Region	Estimated index	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	110.2%	2.1%	1.8%	5.1%
Pref Dalian	110.7%	3.2%	2.9%	6.0%
Pref Fushun	111.3%	3.0%	2.7%	6.1%
Pref Benxi	112.0%	3.1%	2.8%	6.4%
Pref Dandong	109.8%	3.2%	2.9%	6.6%
Pref Jinzhou	107.5%	2.6%	2.4%	6.9%
Pref Yingkou	114.3%	5.2%	4.6%	7.6%
Pref Fuxin	109.0%	3.9%	3.6%	6.9%
Pref Liaoyang	106.8%	2.5%	2.4%	6.0%
Pref Tieling	108.7%	2.6%	2.4%	6.3%
Pref Chaoyang	109.0%	2.8%	2.6%	7.2%
Prov "Liaoning"	109.7%	0.9%	0.9%	1.9%

*) Recall the comments on page 6 concerning possible data deficiencies.

Table 11. Index of change in Crude Birth Rate between 1988 and 1989.				
Region	Estimated index	Absolute margin of sampling error (estimated)	Relative margin of error (estimated)	"Guestimated" relative margin of error
Pref Shengyang	85.7%*	19.5%	22.7%	24.1%
Pref Dalian	78.7%*	19.3%	24.5%	23.0%
Pref Fushun	71.2%*	14.5%	20.4%	22.3%
Pref Benxi	65.2%*	15.6%	23.9%	26.6%
Pref Dandong	89.9%*	19.0%	21.2%	23.6%
Pref Jinzhou	72.6%*	22.5%	31.1%	27.5%
Pref Yingkou	78.8%*	18.7%	23.7%	22.0%
Pref Fuxin	55.2%*	13.6%	24.6%	26.9%
Pref Liaoyang	88.4%*	23.4%	26.5%	25.5%
Pref Tieling	83.6%*	18.1%	21.7%	24.9%
Pref Chaoyang	97.1%*	26.2%	27.0%	23.9%
Prov "Liaoning"	80.3%*	6.6%	8.2%	7.3%

*) Recall the comments on page 6 concerning possible data deficiencies.

4.3 Discussion

The main interest of the numerical findings which are presented in the above tables is that they provide **information on the magnitude of the sampling errors** for estimates of prefecture characteristics as well as of province characteristics. To facilitate overview, we give a compilation in Table 12 below, where average relative margins of errors for prefectures are presented together with the corresponding relative margins of error for the province.

Table 12. Compilation of relative margins of sampling error.

Estimated characteristic	1988: Average relative margin of sampling error in the prefectures.	1988: Relative margin of sampling error for the province "Liaoning".	1989: Average relative margin of sampling error in the prefectures.	1989: Relative margin of sampling error for the province "Liaoning".
Number of eligible women	19.2%	6.2%	19.3%	6.2%
Number of eligible women who use contraceptives	19.4%	6.5%	19.5%	6.5%
Number of eligible women who use IUD	20.0%	6.6%	19.7%	6.5%
Number of children born	27.1%	9.3%	28.5%	9.3%
Contraceptive prevalence among eligible women	5.1%	1.9%	4.7%	1.8%
Proportion IUD users among eligible women	11.3%	3.6%	10.4%	3.4%
Crude Birth Rate	17.6%	6.0%	19.6%	6.4%
General Fertility Rate	18.2%	6.3%	20.1%	6.6%

	Average relative margin of sampling error in the prefectures.	Relative margin of sampling error for the province.
Index of change in proportion IUD users	2.8%	0.9%
Index of change in Crude Birth Rate	24.3%	8.2%

As regards the relation between the relative margins of sampling errors for the same characteristic on prefecture level respectively on province level, the following rule of thumb can be advocated, and arguments for it are given in Section 5.3.

$$\textit{Relative margin of error for an estimated characteristic for a province} \approx$$

$$\approx \frac{\textit{Relative margin of error for the characteristic for a prefecture}}{\sqrt{\textit{Number of prefectures in the province}}}. \qquad (9)$$

""Liaoning" contains 11 prefectures, and $\sqrt{11} \approx 3.3$. It is readily checked that the empirical findings comply quite well with (9).

The material in the Tables 2 - 11 also allows judgement of intra cluster correlations.

- Table 7 shows that high intra-cluster correlation prevails for "IUD use". The "guestimates" of the margins of sampling error are "optimistic" with a factor 2 to 3. Table 6 shows that the situation is similar for contraceptive prevalence in general, but less pronounced.

- Tables 8, 9 and 11 show that intra-cluster correlations as well as correlations over time (cf. Section 5.2.3) for births are practically negligible. "Guestimated" sampling errors for birth rates and change in birth rate lie quite close to the actual values.

- The discrepancies between "guestimated" and actual margins of sampling error which are exhibited in Tables 2 - 5 can, presumably, be ascribed to a combination of intra cluster correlations and the fact that population sizes vary considerably between VGs.

- Table 10 exhibits "guestimated" values which are too "pessimistic". This phenomenon depends, presumably, on strong temporal correlations together with the fact that the MIS VG-samples act as "panels" over time (and panels improve precision for estimates of change). As stated in Section 5.2.3, the "guestimates" are based on the assumption of no temporal correlations. For births it is reasonable to believe that also temporal correlations are low, and Table 11 shows that the "guestimates" work quite well for estimation of the sampling errors for change in birth rates.

A more general impression from the empirical findings is as follows. Even if some of the actual sampling error estimates are larger than indicated by previous "guestimates", the estimation precision for proportions and rates for prefectures are within "reasonable" limits (having a relative margin of error of some 5-15%). However, judgement of estimation precisions along the scale "good - bad" is a complex task, which involves consideration of many aspects; the decisions to be based on the statistics, budgetary matters, etc., to mention some of the more important ones.

5. Computation formulas

In this section we shall specify the formulas that were used in the computation of the numerical findings in the Tables 2 - 11. We start with specifications of the point estimators, which are more formal than those given in Section 3.

5.1 Formal specification of the point estimators

5.1.1 The basic types of statistical characteristics for the MIS surveys

The following notation is used for prefectures as well as provinces. Let

D denote a **group of interest** (e.g. eligible women),

E denote an **attribute group** (e.g. sterilized persons),

$a(D)$ = the **number** of individuals in the group D, (10)

$a(E \cap D)$ = the **number** of attribute-E individuals in group D, (11)

$$p(E;D) = \frac{a(E\cap D)}{a(D)} = \textit{the } \textbf{\textit{proportion}} \textit{ of attribute-E individuals in group } D\,. \tag{12}$$

With the above exemplification, p(E;D) stands for the proportion of "sterilized, eligible women". Let

H denote some **event** during a specified period (e.g. a child birth during 1989),

$a(H\cap D)$ = the **number** of H-events which relate to group D , (13)

$$f(E;D) = \frac{a(H\cap D)}{a(D)} = \textit{the } \textbf{\textit{rate/frequency}} \textit{ of H-events in the group } D\,. \tag{14}$$

Indices of change are denoted as follows, where 0 and 1 refer to different time periods;

$$ch[p(E;D)] = \frac{p(E;D)_1}{p(E;D)_0} \qquad ch[f(H;D)] = \frac{f(H;D)_1}{f(H;D)_0}\,, \tag{15}$$

5.1.2 Point estimation for prefectures

Consider a fixed prefecture, and set;

M = total number of VGs (= village groups) in the prefecture, (16)

m = number of sampled VG:s in the prefecture. (17)

The letter ν is used to index the sampled VGs within a prefecture and, hence, ν runs over $\nu=1,2,...,m$. Observed total numbers in sampled VGs are denoted by capital A:s;

$A_\nu(D)$ = number of individuals in group D in sample VG no ν, $\nu=1,2,...,m$, (18)

$A_\nu(E\cap D)$ = number of attribute-E individuals in group D in sample VG no ν, $\nu=1,2,...,m$. (19)

$A_\nu(H\cap D)$ = number of H-events relating to group D in sample VG no ν, $\nu=1,2,...,m$. (20)

In accordance with (1), totals numbers, proportions, rates and indices of change for the prefecture are estimated according to (21) - (24) below, where "hats" (^) indicate estimators. The estimators in (21) - (24) are formal versions of those which were stated in Section 3.

$$\hat{a}(D) = \frac{M}{m}\cdot\sum_{\nu=1}^{m} A_\nu(D)\,, \tag{21}$$

$$\hat{a}(E\cap D) = \frac{M}{m}\cdot\sum_{\nu=1}^{m} A_\nu(E\cap D)\,, \quad \hat{a}(H\cap D) = \frac{M}{m}\cdot\sum_{\nu=1}^{m} A_\nu(H\cap D)\,, \tag{22}$$

$$\hat{p}(E;D) = \frac{\hat{a}(E\cap D)}{\hat{a}(D)}, \qquad \hat{f}(H;D) = \frac{\hat{a}(H\cap D)}{\hat{a}(D)}, \tag{23}$$

$$\hat{ch}[p(E;D)] = \frac{\hat{p}(E;D)_1}{\hat{p}(E;D)_0} \qquad \hat{ch}[f(H;D)] = \frac{\hat{f}(H;D)_1}{\hat{f}(H;D)_0}, \tag{24}$$

5.1.3 Point estimation for provinces

As stated before, the sample from a province is viewed as a stratified sample with the prefectures as sampling strata. Let

K = the number of sampling strata (= prefectures) in the province. (25)

The letter k is used to index prefectures within a province and, hence, k runs over k=1,2, ...,K. M_k and m_k denote the total number of VGs respectively the VG sample size in prefecture no k. In the province context we let a(D), a(E∩D) and a(H∩D) denote the numbers **within the province**, while a_k(D), a_k(E∩D) and a_k(H∩D) denote the corresponding numbers for prefecture no k, k=1,2,...,K. Similarly, p(E;D) and f(H;D) denote proportions and rates/frequencies in the province, while the corresponding quantities for prefecture no k are denoted by p_k(E;D) and f_k(H;D). As before ν indexes sampled VGs.

Total numbers, proportions, rates and indices of change for provinces are estimated as stated in (26) - (30) below, and again the formulas only formalize the estimators which were stated in Section 3;

$$\hat{a}(D) = \sum_{k=1}^{K} \hat{a}_k(D) = \sum_{k=1}^{K} \frac{M_k}{m_k} \cdot \sum_{\nu=1}^{m_k} A_{k\nu}(D), \tag{26}$$

$$\hat{a}(E\cap D) = \sum_{k=1}^{K} \hat{a}_k(E\cap D) = \sum_{k=1}^{K} \frac{M_k}{m_k} \cdot \sum_{\nu=1}^{m_k} A_{k\nu}(E\cap D), \tag{27}$$

$$\hat{a}(H\cap D) = \sum_{k=1}^{K} \hat{a}_k(H\cap D) = \sum_{k=1}^{K} \frac{M_k}{m_k} \cdot \sum_{\nu=1}^{m_k} A_{k\nu}(H\cap D), \tag{28}$$

$$\hat{p}(E;D) = \frac{\hat{a}(E\cap D)}{\hat{a}(D)}, \qquad \hat{f}(H;D) = \frac{\hat{a}(H\cap D)}{\hat{a}(D)}, \tag{29}$$

$$\hat{ch}[p(E;D)] = \frac{\hat{p}(E;D)_1}{\hat{p}(E;D)_0} \qquad \hat{ch}[f(H;D)] = \frac{\hat{f}(H;D)_1}{\hat{f}(H;D)_0}, \tag{30}$$

5.2 Estimation of margins of sampling errors

5.2.1 Variance estimation for estimates of prefecture characteristics

The variance estimators which were inserted in (6) were computed by the variance formulas which emanate from the following assumption (31) about the probabilistic structure of the sample. Note that (31) complies with (1), and means a specialization of it.

The VG sample from a prefecture is a simple random sample of VGs. (31)

The assumption (31) can of course be questioned on the ground that the samples of VGs from the prefectures were drawn by systematic sampling. As is well known, the issue of variance estimation for systematic sampling is complex. However, under the assumption that variable values appear as fairly randomly ordered in the sampling frame, systematic sampling and simple random sampling are probabilistically equivalent. We do not have good enough knowledge of the sampling frame ordering to make sure judgement on this point, but we believe that the simple random sampling assumption is sufficiently realistic. We therefore resort on the common practice to base the variance estimations for systematic sampling on the assumption (31). Moreover, the figures in Table 1 justify the following simplification;

Sampling fractions are so small that finite population correction factors can be disregarded. (32)

Thereby the variance estimation problem is brought back on the situation of estimating variances for estimators at cluster sampling with simple random sampling of the clusters. This problem is treated in most books on sampling theory, and a possible source is Cochran (1977), Chapter 9. However, even if we rely on well-known results we list, for the sake of completeness, the precise versions of the computation formulas which were employed. The following notation is used. For a set of sample values $X=(X_1, X_2, ..., X_m)$, the corresponding ("ordinary") **sample variance** is denoted as follows, where $\bar{X}$ denotes **sample mean**;

$$S^2(X) = \frac{1}{m-1}\cdot\sum_{v=1}^{m}(X_v - \bar{X})^2 . \tag{33}$$

Estimation of variances for the estimates of total number within the prefecture, a(D), a(E∩D) and a(H∩D), was carried out along the following lines;

$$\hat{V}[\hat{a}] = \hat{V}\left[\frac{M}{m}\cdot\sum_{v=1}^{m}A_v\right] = \frac{M^2}{m}\cdot S^2(A) . \tag{34}$$

The estimators in (29) of proportions and rates are of ratio type. The usual "Taylor approximation formula" for the variance of a ratio variable leads to the following variance estimation formula in the p-case, and the f-case was treated quite analogously;

$$\hat{V}[\hat{p}(E;D)] = \hat{V}\left[p(E;D)\cdot\frac{M}{m}\cdot\sum_{v=1}^{m}\left(\frac{A_v(E\cap D)}{a(E;D)}-\frac{A_v(D)}{a(D)}\right)\right]=$$

$$= [\hat{p}(E;D)]^2\cdot\frac{1}{m}\cdot S^2\left[\frac{A(E\cap D)}{\hat{a}(E\cap D)}-\frac{A(D)}{\hat{a}(D)}\right]. \tag{35}$$

The estimators of change in (30) are of double ratio type. By applying Taylor approximation for this kind of estimator one arrives at the following variance estimation formula (36) in the p-case. Again, the f-case was treated quite analogously.

$$\hat{V}(\hat{ch}[p(E;D)]) = (\hat{ch})^2\cdot\frac{M^2}{m}\cdot S^2\left[\frac{A_v(E\cap D)_1}{\hat{a}(E\cap D)_1}-\frac{A_v(D)_1}{\hat{a}(D)_1}-\frac{A_v(E\cap D)_0}{\hat{a}(E\cap D)_0}+\frac{A_v(D)_0}{\hat{a}(D)_0}\right]. \tag{36}$$

In full the formula (35) appears as follows, and (36) can be written quite analogously;

$$\hat{V}[\hat{p}(E;D)] = [\hat{p}(D)]^2\cdot\frac{1}{m\cdot(m-1)}\cdot\sum_{v=1}^{m}\left[\frac{A_v(E\cap D)}{\hat{a}(E\cap D)}-\frac{A_v(D)}{\hat{a}(D)}\right]^2. \tag{37}$$

Note the following aspect on (37). When computing the variance S^2 in (35) one only gets a "sum of squares" contribution and no "square of sum" contribution, due to the following relation;

$$\sum_{v=1}^{m}\left(\frac{A_v(E\cap D)}{\hat{a}(E\cap D)}-\frac{A_v(D)}{\hat{a}(D)}\right)=0. \tag{38}$$

5.2.2 Variance estimation for estimates of province characteristics

Here we let A_k denote the collection of total numbers for the sampled VGs in prefecture no k, i.e. $A_k = (A_{kv}\ ;\ v = 1,2,...,m_k)$. By employing (31) together with the fact that

The VG samples from the different prefectures are drawn independently of each other, (39)

estimators of the variances for estimators of province characteristics are obtained by well-known generalizations to the stratified sampling case of the previous results for prefectures. Again, Chapter 9 in Cochran's book is a possible source. The formulas are listed below, and the "missing" formulas are quite analogous to the listed ones.

$$\hat{V}[\hat{a}] = \sum_{k=1}^{K}\hat{V}\left[\frac{M_k}{m_k}\cdot\sum_{v=1}^{m_k}A_{kv}\right]=\sum_{k=1}^{K}\frac{M_k^2}{m_k}\cdot S^2(A_k). \tag{40}$$

$$\hat{V}\left[\hat{p}(E;D)\right] = [\hat{p}(E;D)]^2 \cdot \sum_{k=1}^{m_k} \frac{M_k^2}{m_k} \cdot S^2 \left[\frac{A_k(E \cap D)}{\hat{a}(E \cap D)} - \frac{A_k(D)}{\hat{a}(D)} \right]. \tag{41}$$

$$\hat{V}\left[\hat{ch}[p(E;D)]\right] =$$

$$= (\hat{ch}[p(E;D)])^2 \cdot \sum_{k=1}^{m} \frac{M_k^2}{m_k} \cdot S^2 \left[\frac{A_k(E \cap D)_1}{\hat{a}(E \cap D)_1} - \frac{A_k(D)_1}{\hat{a}(D)_1} - \frac{A_k(E \cap D)_0}{\hat{a}(E \cap D)_0} + \frac{A_k(D)_0}{\hat{a}(D)_0} \right]. \tag{42}$$

Note the following. In formulas (41) and (42), $\hat{a}(D)$ and $\hat{a}(E \cap D)$ are the province estimates. Moreover, in this case there are no counterparts to (38), and accordingly the variances in (41) and (42) can not be computed by "simplified variance formulas" as in (37), they have to be computed by "full" variance formulas.

5.2.3 Computation of the "guestimates" of the margins of sampling error

The "guestimate" formulas refer to the case specified by assumption (8), i.e. the case with simple random samples of individuals from the prefectures, with sample sizes as listed in Table 1. Also here we disregard finite population correction factors. For a simple random sample of size n from a population of size N the following formulas hold, and again we refer to e.g. the book by Cochran, notably Chapters 2 and 3. "Missing" formulas are parallel to the listed ones. RME = (theoretical) relative margin of error.

$$RME\left[\hat{a}(D)\right] = \frac{2}{\sqrt{n}} \cdot \sqrt{\frac{N}{a(D)} - 1} \ , \tag{43}$$

$$RME\left[\hat{p}(E;D)\right] = \frac{2}{\sqrt{n \cdot [a(D)/N]}} \cdot \sqrt{\frac{1}{p(E;D)} - 1} \ , \tag{44}$$

$$RME\left[\hat{f}(H;D)\right] = \frac{2}{\sqrt{n \cdot [a(D)/N]}} \cdot \sqrt{\frac{1}{f(H;D)} - 1} \ , \tag{45}$$

To get a "guestimate" formula for the RME of indices of change, one also has to make assumptions on the correlation between observations for different years. On this point we employ the assumption that attributes and events occur independently of each other for compared years. (The realism of this assumption can of course be questioned, as can assumption (9).) Then the following holds;

$$RME\left[\hat{ch}[f(H;D)]\right] \approx \sqrt{\left(RME\left[\hat{f}(H;D)_1\right]\right)^2 + \left(RME\left[\hat{f}(H;D)_0\right]\right)^2} \tag{46}$$

The numerical values for the "guestimates" which are reported in Tables 2 -11, were derived by substituting the theoretical quantities in (43) - (46) (N, a(D), p(E;D), etc.) by the corresponding estimated values.

5.3 Comments on the formula (9)

We shall here give some heuristic justification for the rule of thumb which was stated in (9). The basic ingredient in the arguments is that estimates from different prefectures are independent of each other. The technical argumentation differs somewhat for different characteristics, but we confine ourselves to the case with proportions. Upon some thought it is realized that the estimator in (29) can be approximately expressed as follows;

$$\hat{p}(E;D) \approx \frac{1}{K} \cdot \sum_{k=1}^{K} \hat{p}_k(E;D) . \tag{47}$$

Now, as the samples from the different strata are drawn independently of each other (see (39)), the sum in (47) is a sum of independent estimates. Hence, by well-known rules for variances;

$$V\left[\hat{p}(E;D)\right] \approx \frac{1}{K^2} \cdot \sum_{k=1}^{K} V[\hat{p}_k(E;D)] . \tag{48}$$

Under the premise that conditions in the prefectures are "fairly similar", the terms in (48) are "fairly equal" and (48) simplifies as follows;

$$V\left[\hat{p}(E;D)\right] = \frac{1}{K} \cdot (\textit{the value of } V[\hat{p}_k(E;D)] \textit{ for a "typical" prefecture}) . \tag{49}$$

By taking square roots of both sides in (49) the formula (9) is obtained.

References

COCHRAN, W.G. (1977). *Sampling Techniques, Third Edition.* Wiley, New York.

NYGREN, O. Discussion and Negotiations on the New Agreement's Framework - Sample Size and the Computer System's Design. SCB/ICO, KINASTAT 1990:1.

POLFELDT, T. On the Sample Design of a Data Base for Family Planning Monitoring. SCB/ICO, KINASTAT 1988:2.

R & D Reports är en för U/ADB och U/STM gemensam publikationsserie, som fr o m 1988-01-01 ersätter de tidigare "gula" och "gröna" serierna. I serien ingår även **Abstracts** (sammanfattning av metodrapporter från SCB).

R & D Reports Statistics Sweden are published by the Department of Research & Development within Statistics Sweden. Reports dealing with statistical methods have green (grön) covers. Reports dealing with EDP methods have yellow (gul) covers. In addition, abstracts are published three times a year (light brown /beige/ covers).

Reports published during 1991:

1991:1 (grön)	Computing elementary aggregates in the Swedish consumer price index **(Jörgen Dalén)**
1991:2 (grön)	Översikt av estimatorer för stora och små redovisningsgrupper **(Sixten Lundström)**
1991:3 (grön)	Interacting nonresponse and response errors **(Håkan L Lindström)**
1991:4 (grön)	Effects of nonresponse on survey estimates in the analysis of competing exponential risks **(Ingrid Lyberg)**

Kvarvarande **beige** och **gröna** exemplar av ovanstående promemorior kan rekvireras från Inga-Lill Pettersson, U/LEDN, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 49 56.

Kvarvarande **gula** exemplar kan rekvireras från Ingvar Andersson, U/ADB, SCB, 115 81 STOCKHOLM, eller per telefon 08-783 41 47.