Third Baltic–Nordic Conference in Survey Statistics

June 13-17, 2011

Norrfällsviken, Sweden

PROCEEDINGS

Baltic-Nordic-Ukrainian Network of Survey Statistics
Universities of Stockholm, Umeå and Helsinki

Third Baltic–Nordic Conference in Survey Statistics
June 13-17, 2011, Norrfällsviken, Sweden

Programme Committee

Jan Bjørnstad, Statistics Norway
Eva Elvers, Statistics Sweden
Danute Krapavickaite, Institute of Mathematics and Informatics, Vilnius Gediminas Technical University and Statistics Lithuania
Gunnar Kulldorff, University of Umeå
Janis Lapins, Bank of Latvia
Risto Lehtonen, University of Helsinki
Peter Linde, Statistics Denmark
Aleksandras Plikusas, Institute of Mathematics and Informatics
Daniel Thorburn, University of Stockholm (Chair)
Imbi Traat, University of Tartu
Olga Vasylyk, National Taras Shevchenko University of Kyiv

Organizing Committee

Eva Elvers, Statistics Sweden (until September 2010)
Gunnar Kulldorff, University of Umeå (Chair)
Risto Lehtonen, University of Helsinki
Joakim Malmdin, Statistics Sweden (from October 2010 on) (Secretary)
Daniel Thorburn, University of Stockholm
Imbi Traat, University of Tartu

Organisers

Baltic-Nordic-Ukrainian Network of Survey Statistics
University of Umeå
Stockholm University
University of Helsinki

Sponsors

The Swedish Institute
The Swedish Research Council
The Kempe Foundation
The Nordic Council of Ministers
The International Association of Survey Statisticians (IASS)

FOREWORD

This is the third Baltic-Nordic conference and it is a part of a series of workshops and conferences, initiated in 1997 by Professor Gunnar Kulldorff and organized since then annually in different Baltic and Nordic countries and from 2008 also in Ukraine. The main organizer has been the Baltic-Nordic Network in Survey Sampling which in recent years has been enlarged to a Baltic-Nordic-Ukrainian network. It contains of people from university departments, national statistical institutes and statistical societies.

The globalisation and the development of internet, registers and computing facilities have contributed to making survey sampling into a growing and vigorous science. The key-note speakers reflect both developments within traditional sampling theory, modern techniques like nonparametric sampling and sampling in non-standard situations. The present conference mirrors this development with contributions on such diverse matters as multinational studies and computer packages for editing. In the present conference there are more than 60 participants from 18 countries. This programme contains almost 50 papers. They represents a broad spectrum of what is going on in sampling survey research today.. It covers topics from mathematical treatments to behavioural sciences and from deep theoretical questions to problems which have been encountered in everyday work as a statistician.

Authors of conference papers are encouraged to submit manuscripts for publication in a special issue of Statistics in Transition Journal. We thank Editor-in-Chief Professor Włodzimierz Okrasa for offering pages of the journal for this purpose.

I would also like to thank the other members of the programme and the organizing committees and all others who have contributed to it. In particular I want to mention Maria Valaste for her work with the website and all other people who have contributed in different ways, Finally I would like to thank The Swedish Institute, The Swedish Research Council, The Kempe Foundation, The Nordic Council of Ministers and The International Association of Survey Statisticians (IASS) for their generous support.

I wish you all a enjoyable stay and an inspiring conference in Norrfällsviken and I hope that you will leave here with many nice memories and also new knowledge and insights.

Stockholm, June 2011

Daniel Thorburn

Contents

CALIBRATION, BALANCED SAMPLING WITH APPLICATIONS TO NON-RESPONSE

Jean-Claude DEVILLE
ENSAI/Crest
Laboratoire de Statistique d'Enquête

GENERALITIES

1-Universe, label, sample and sample design, inclusion probabilities

2- Unbiased estimation for a total

3-Practical Sampling

→Decomposition of a sampling plan: Stratification and multi-stage designs

→Units sampling plans

4-Non-linear Estimation: Plug-in Principle (or substitution)

→Examples of non-linear indicators to be estimated

→Plug-in principle for point estimation

→Linearization for variance approximation with examples

→Linearization variance estimation

5-Representativeness à la Hajek

BALANCED SAMPLING

1-Entropy and Poisson Sampling

2-Variance, Variance Approximation and Variance Estimation and Extensions

2-1Necessity of an approximation

2-2Properties of a good approximation

3-Exact or approximate balance?

4-The Cube Method

4-1 The Cube set-up

4-2 Flight (iterative) phase

4-3 Landing (round-up) phase

5-Variance and Variance Estimation

5-1 How to get maximum entropy?

5-2 Ordering and the quick sequential algorithm

CALIBRATION MORE OR LESS GENERALIZED

1-TRADITION:ratio,poststratification,regression

2-Calibration Principle :Goal and examples

→Examples

→Principle

→Some general properties

→A way to do that or playing with the weights

3-Generalized Calibration:Estimation knowing totals

3-1The case of a linear link function

3-2 Properties: Bias, variance

3-3 Practical way (CALMAR II software) and applications

3-4 Some ideas on the numerical aspects

4-Standard Calibration and Applications

4-1→Choice of the instruments; choice of the link function
4-2→Determination using a distance function between the weights
4-3→Raking-ratio revisited
4-4→Calibration on inexact (unbiased) data ; simultaneous calibration of several surveys

5- Two-phases Sampling and Calibration

5-1 Two phases sampling: estimator , variance and variance estimation
5-2Auxiliary information and calibration in two phases sampling
5-3 Two phases sampling as a model for the response mechanism (an introduction)

APPLICATIONS TO NON-RESPONSE

A_-Generalities

1-Non-response is an affair of principles!!

2-Pattern of Non-response and Correction Methods

2-1 Total (unit) and partial (item) non-response
2-2 Goals of the corrections

3-Basic methods:weighting or imputing?

3-1Outline of the methods
a-Weighting
b-Imputing
3-2What Method for What Use?

B-Weighting for non-response and generalized calibration

1-Model

2-Examples: Parametric and non-parametric models

3- **Estimation of the parameters**

3-1 Incidence of the use of estimated parameters for estimation
3-2 How to estimate the parameters of the response mechanism?

4-Variance and estimating equations

5-Examples

C-Deterministic imputation, prediction and calibration

1-The estimator and the superpopulation model

2-Derivation from a superpopulation model

3-Estimating equations for parameters and variance

4-Choice of the instruments

4-1Standard considerations
4-2 Trick avoiding to compute explicitly the response weights

5-Examples : ratio imputation, regression imputation

6-Summary

D-Random imputation and balanced sampling

0-Imputation 'Weighting Like'

1-Generalities

1-1Some new problems
1-2-Quick inventory of methods

2-Formalization and Estimation Theory

2-1The extra-variance
2-2 Exercise (or example!)
2-3 General recipes for the extra variance

3-Applications, Examples and Concluding Remarks

3-1- Imputation of a (0-1) variable and balanced sampling
3-2-Qualitative variable and balanced sampling of cells
REMARK: Is random imputation really mandatory?
3-3-Quantitative variable and balanced sampling of residuals
1-Gaussian model.
2- I like the following technique
3- Use the Cube.
REMARK: Is random imputation really mandatory?
3-4-Summary and final remarks

Nonparametric regression in inference for finite populations

M. Giovanna Ranalli[1]

In recent years, the growing availability of auxiliary information in inference for finite populations has motivated the development of estimators that incorporate such information that are alternative to the unbiased Horvitz-Thompson estimator. In a design-based estimation framework, such estimators aim at increasing precision at the expenses of the introduction of a little bias. In the past decades, two main approaches have been used to this end: generalized regression (GREG) estimation and calibration (CAL) estimation. Each of these two approaches provides a whole class of estimators that share some similarities, but that can also be quite different. The former is explicitly assisted by a superpopulation model. Estimators that belong to the latter, on the other hand, are developed without any explicit reference to an assisting model, but it is well-known that linear models that can be associated with some of them.

In this set of talks we will first review the extension of the classical GREG and CAL estimators to a wider set of assisting parametric models (non-linear, generalized linear and mixed models). We will then look closely at estimators in these two frameworks that are assisted by or associated to nonparametric regression models. These are models in which the relationship between the variable of interest and one (or more) auxiliary variable(s) does not have a pre-specified parametric form, but is left undefined and learnt from the data. In particular, we will consider GREG and CAL estimators that use the following nonparametric regression techniques to approximate such relationship(s): Kernel and Local Polynomials, Generalized Additive Models, penalized Splines, Generalized Additive Mixed Models, Neural Networks. Most, but not all, of the estimators considered require complete auxiliary information (i.e. the value of a set of auxiliary variables has to be known for each unit in the population).

The use of nonparametric regression techniques has also been introduced in surveys in those fields in which statistical models play an important role. We will then review the use of nonparametric regression models for treatment of nonresponse and measurement error, and, in a model-dependent framework, for small area estimation.

An illustration of the methods discussed will be provided on data from an environmental survey of lakes in the North-Eastern US conducted by EPA, from the Labor Force Survey conducted by ISTAT, and from the Italian Survey of Households' Income and Wealth conducted by the Bank of Italy.

References

Breidt, F.J. and J.D. Opsomer (2009). Nonparametric and semiparametric estimation in complex surveys, in *Handbook of Statistics - Sample Surveys: Inference and Analysis,* Vol. 29B, D. Pfeffermann and C.R. Rao (editors), The Netherlands: North-Holland, 103-120.

Da Silva, D. N. and J. D. Opsomer (2006). A kernel smoothing method of adjusting for unit non-response in sample surveys. *The Canadian Journal of Statistics* 34(4), 563–579.

Montanari, G. E. and M. G. Ranalli (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association* 100 (472), 1429–1442.

Opsomer, J. D. and C. P. Miller (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *Journal of Nonparametric Statistics* 17-5, 593–611.

Opsomer, J. D., G. Claeskens, M. G. Ranalli, G. Kauermann, and F. J. Breidt (2008). Nonparametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 70 (1), 265–286.

Särndal, C.E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* 33(2), 99–119.

Wu, C. and Y. Luan, (2003). Optimal calibration estimators under two-phase sampling. *Journal of Official Statistics*, 19, 119–131.

Wu, C. and R. R. Sitter (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association* 96, 185–193.

[1] Dept. of Economics, Finance and Statistics, University of Perugia, Italy - giovanna.ranalli@stat.unipg.it

ADAPTIVE SAMPLING

Steven Thompson[1]

Adaptive sampling includes designs for sampling populations that are elusive, rare, uneven, or hard to detect. Some of the work is focused on sampling in networks and some on sampling in spatial settings and also where sampling units are in motion.

A lot of the spatial sampling work has been motivated by problems in environmental studies, including ecological surveys of rare, clustered animal and plant species. The network sampling work has been motivated by problems in studies of hidden human populations such as those at high risk for HIV/AIDS, including injecting drug users and commercial sex workers.

[1] Simon Fraser University 8888 University Drive, Burnaby, BC V5A 1S6, CANADA

BANOCOSS, Norrfällsviken, Sweden, June 13-17, 2011

Extended Sampford sampling, balanced Pareto sampling, and sampling with prescribed second-order inclusion probabilities

Lennart Bondesson

Dept. of Math. & Math. Statistics, Umeå Univ., SE-90187 Umeå, Sweden
e-mail: Lennart.Bondesson@math.umu.se

ABSTRACT

This talk on unequal probability sampling has three parts.

i) Sampford sampling is extended to the case that the inclusion probabilities do not sum to an integer. In terms of sampling indicators, the sample will be of the form $[0, 1, 1, 0, 1, a, 0, 1, 0, 0]$, where $0 < a < 1$. For exactly one randomly chosen unit the sampling outcome is left undecided. The method has several applications.

ii) Balanced (or restricted) Pareto sampling is generalization of Pareto sampling to the case that there are several linear restrictions on the sample. Like for ordinary Pareto sampling, adjustment is needed to get the desired inclusion probabilities.

iii) The third part of the talk treats the old unsolved problem of sampling with prescribed second-order inclusion probabilities. Different solutions are presented. Second-order conditional Poisson sampling is one possibility.

REFERENCES

Bondesson, L. (2010). Conditional and restricted Pareto sampling: two new methods for unequal probability sampling. *Scand. J. Statist.* **37**, 514-530.

Bondesson, L. & Grafström, A. (2011). An extension of Sampfords method for unequal probability sampling. *Scand. J. Statist.* **38**, 377-392.

Bondesson, L. (2011). Sampling with prescribed second-order inclusion probabilities. Manuscript, 34 pages.

Small Area Estimation for Poverty Indicators

Risto Lehtonen[1], Ari Veijanen[2], Mikko Myrskylä[3] and Maria Valaste[4]

Abstract

The paper discusses some aspects of the estimation of indicators on monetary poverty (so-called Laeken indicators of the EU) for population subgroups or domains and small areas. Research was conducted in 2008-2011 in the context of the AMELI project (Advanced Methodology for European Laeken Indicators; Lehtonen *et al*. 2011), which was supported by European Commission funding from the Seventh Framework Programme for Research.

The set of Laeken indicators we investigated in AMELI includes at-risk-of poverty rate (poverty rate for short), relative median at-risk-of poverty gap, quintile share ratio and the Gini coefficient. Our numerical results presented here concentrate on poverty rate and quintile share ratio.

In poverty rate estimation we deal with a binary study variable indicating whether a person is in poverty or not, based on definitions adopted in the EU. We used design-based logistic generalized regression and model calibration estimators and model-based empirical best predictor type estimators. The estimators are indirect aiming at borrowing strength from similar areas. In construction of the estimators we used logistic mixed models with area-specific random intercepts. Unit-level auxiliary data were incorporated in the estimation procedure. Horvitz-Thompson and related design-based direct estimators represent the reference estimators. Our goal is to introduce estimators that are more accurate than conventional direct estimators.

In quintile share ratio estimation we deal with 20% and 80% quintiles of the cumulative distribution function of the underlying equivalized income variable. Design-based direct, model-based indirect and composite estimators were investigated. Linear mixed models with area-specific random intercepts were fitted to the (transformed) income variable at the unit level for the construction of prediction and expanded prediction type estimators. Composite estimators were constructed as a linear combination of a direct estimator and an indirect expanded prediction estimator.

The relative performance of estimators (design bias and accuracy) was examined with design-based simulation experiments. Equal and unequal probability sampling designs were covered. We also examined the properties of the estimators under outlier contamination. In the experiments we used unit-level data obtained from statistical registers maintained by Statistics Finland and a more extensive Amelia population constructed from the EU-wide SILC survey data.

Key words: Poverty indicator, Small area estimation, GREG, Empirical Best Predictor, Composite estimator, Mixed model, Outlier contamination

References

Lehtonen R., Veijanen A., Myrskylä M. and Valaste M. (2011). Small Area Estimation of Indicators on Poverty and Social Exclusion. AMELI Deliverable 2.2 of Work Package 2. (Unpublished manuscript of May 2011)

[1] University of Helsinki, risto.lehtonen@helsinki.fi
[2] Statistics Finland, ari.veijanen@stat.fi
[3] Max Planck Institute for Demographic Research, Germany, myrskyla@demogr.mpg.de
[4] The Social Insurance Institution of Finland, maria.valaste@kela.fi

SELECTIVE DATA EDITING

Anders Norberg[1]

Editing is an activity of detecting, resolving and understanding errors in data and produced statistics.

Errors in raw data delivered by respondents to the statistical agency are typically measurement errors and non-response. For technical reasons errors can be introduced in data transmission. The production of statistics is a mixture of sub-processes such as joining with register data, coding, imputation, estimation, seasonal adjustment and analysis. In these sub-processes there is a risk of introducing errors. There are a series of editing-activities of various kinds, starting at the stage of data collection with built in checking rules in electronic questionnaires and ending with control of tables and diagrams in statistical publications.

Editing is a critical and expensive part of the statistical operation. Many national statistical offices recognize that they devote about one third of the resources for production of enterprise statistics to editing. The most expensive editing activity is the "micro-editing" which is checking for errors in micro data when they are stored electronically in input databases. There is a potential for big cut of costs if editing can be more efficient. This potential exist as many editing systems today are designed to find all errors in data. If the objective of doing a survey is the set of pre-specified statistical tables adopted for the users´ needs – and no capability for extra analysis – then the search for errors can be directed to those that have significant impact on the survey output. This is the idea of selective data editing.

The basic properties for a raw data value, tested by selective data editing, are (1) suspicion for being in error, (2) the potential impact on the output statistics conditional on being in error and (3) importance parameter values for the domain and survey variable relative to the output as a whole.

The purpose of selective data editing is to reduce cost for the statistical agency as well as for the respondents, without significant decrease of the quality of the output statistics. The follow-up of data that are flagged for being in error with certainty or with a "suspicion" often require re-contacts with respondents.

This purpose of selective data editing does not clearly support the main objectives of editing and imputation that are beginning to be recognized. These are:

1. identify error sources in order to provide information for future improvements of the survey process;
2. provide information about the quality of the incoming/outgoing data;
3. identify and resolve the most significant errors;
4. when needed, provide complete and consistent (coherent) individual data.

[1] Statistics, Sweden

Survey Statistics : the past, the present, the future

Carl-Erik Särndal
Statistics Sweden
Örebro University

BaNoCoSS
Höga Kusten, Sweden
June 13-17, 2011

2011-05-12

Background of my talk : "The Canada Census Incident"

The Government of Canada made Statistics Canada abandon "the long form" (systematic one-in-five households) for detailed 2011 census information,

and replace it with a voluntary *National Household Survey* (1/3 of all households)

High degree of abstention (nonresponse) expected.

The Canada Census Incident

The Government says : Respect the right of Canadians to refuse to divulge personal information.

The Statisticians complain : Accuracy will suffer.

The Users complain : Time series breaks; unreliable information about many groups in society.

The Canada Census Incident

With the new design,
Canada's population will be shown as richer, better educated, more conservative than it is,

at the expense of accurate information about small and disadvantaged groups.

The Canada Census Incident

Question arising:
Can non-statisticians tell professional statisticians how statistics of national importance should be produced?

Of similar kind :
Can ordinary people tell surgeons how brain surgery, for example, should be carried out?

Questions arising

- Could it be that official statistics production is so elementary, so lacking in established theory that questionable common sense can be allowed to take over ?
- Information of unknown or doubtful accuracy, produced against better knowledge, is that nevertheless *preferable* to high quality information conforming to highest of standards?
- Is official statistics production scientific ? If yes, to what extent ?

The Canada Census Incident

it actually happened – regrettably

Outline of my talk

Five brief comments :

1. Official statistics and scientific principles.
2. *Statistical science* vis-à-vis *official statistics production*.
3. Official statistics: a fragmented field
4. *Survey methodology* vis-à-vis *survey theory*
5. About the future

1. Official statistics and scientific principles.

Main objective of an NSI : Deliver high quality statistics in high demand by users.

The NSI:s – some at least - pride themselves in "scientific principles"

Statistics Sweden : "The statistics produced rely on a scientific foundation"

1. Official statistics and scientific principles.

There is not one unified (comprehensive) theory for official statistics.

Article by Robert Groves (1987) titled:

Survey research is a methodology without a unifying theory

1. Official statistics and scientific principles.

No unifying theory for official statistics.

If statistics production could point to a firm solid theory, enjoying the prestige of recent major scientific breakthroughs,

there would be no room for a Canada Census Incident.

1. Official statistics and scientific principles.

Many observers (in high places) see official statistics production as

a bundle of ***techniques***

with some theory here and there

Statisticians are seen as technicians, "number-crunchers"

Nevertheless they are very important people

1. Official statistics and scientific principles.

NSI:s also recognize the absence of unifying theory.

Statistics Canada (1998) : Survey Methodology is

"a collection of practices, backed by some theory and empirical evaluation, among which practitioners have to make sensible choices in the context of a particular application"

1. Official statistics and scientific principles.

The cited article by Groves (1987) about non-existence of unified theory:

"A theory of surveys would unite social science concepts with the statistical properties of survey estimates" (i.e., accuracy; bias and variance)

2. Statistical science vis-à-vis statistics production

A central idea in ***statistical science***, as taught in many universities :

From a part (the sample), make ***probability*** statements about the whole (the population).

This is ***statistical inference***

Statements - significant differences , confidence intervals - at specified level of ***probability***

2. Statistical science vis-à-vis statistics production

A central question in ***statistical science***:

" How far from the truth; "how close are we"

The theory had its heyday in the 1930's and 1940's.

2. Statistical science vis-à-vis statistics production

Official statistics production gives us "*numbers about the population*"

as opposed to "*inference about the population*" as offered by ***Statistical science***

The usual concepts (confidence statements, etc.) are not operational in official statistics.

There they never say "we are this close to the truth"; they say instead "we do the best we can"

2. Statistical science vis-à-vis statistics production

At this point of my presentation, some listeners start to feel uncomfortable :

"What do you mean we don't make inferences ?"

Reply: Statistics production does use statistical theory,

in various bits and pieces

also profits from theory from other sciences

but it does not make inferences

2. Statistical science vis-à-vis statistics production

Here, I am just asking myself some questions;
sharing with you a perspective on values that we all hold as statisticians.

A scientific view - from inside official statistics

2. Statistical science vis-à-vis statistics production

Franchet and Nanopoulos (1997) article titled *Statistical science and the European statistical system: expectations and perspectives*

"The methodology of official statistics is a notion that has to be distinctly understood from the notion of methodology in mathematical statistics"

"The probabilistic formalism … of mathematical statistics has offered official statistics the necessary framework for its scientific foundation."

2. Statistical science vis-à-vis statistics production

I see a wide gap between

the principles of ***statistical science***
and
the stark reality of today's official ***statistics production***

3. Official statistics production: a fragmented field

Fragmentation (of a field of knowledge) is a concept in philosophy of science.
Two of its aspects :
(a) Competing theories within the field creates divisions
(b) "Piecemeal theory" develops within a field that should be more unified

Term ***Fragmentation*** not derogatory, just descriptive. Reference: ***Science, order, and creativity*** by D. Bohm and F. D. Peat (2000)

David Bohm (1917-1992), quantum physicist

3. Official statistics: a fragmented field

There is *fragmentation* when divisions arise in a more or less arbitrary fashion without any regard for a wider context
Ref: B&P p. 15

3. Official statistics: a fragmented field

A sign of fragmentation is the emergence of separate groups of investigators, held together by common interest in a certain (limited) question.
A group of people get together and work on the same problem, under a trademark name

Official statistics has many examples :
Imputation, Nonresponse weighting, Editing and data cleaning, Small area estimation, and so on

3. Official statistics: a fragmented field

As time goes by, problem areas arise in a science, some become more and more "burning", engender a phase of development.

Theory develops within narrow sub-fields, pieces of theory, specializations inside the broader field, highly specific areas of knowledge, subcultures .

So it is with statistics production: It has come to rely on "a collection of practices, backed by some theory here and there"

3. Official statistics: a fragmented field

Some official statistics subcultures :

In data treatment:

Small area estimation
Nonresponse weighting
Imputation
Editing and data cleaning

In data delivery:

Response burden
Motivating respondents
Confidentiality protection

3. Official statistics: a fragmented field

Official statistics: Active groups, networks, exist in a number of ***narrow specializations***.

Is this good, in the long run ?
Where will it take us ?

3. Official statistics: a fragmented field

"Long range connections between the ideas is of crucial importance in the continued development of a field, and they cannot be dealt with in terms of narrow specializations" (Ref: B&P p. 71)

Regrettable, but for official statistics production, is there an alternative ?

4. Survey theory versus Survey methodology

In a history of science perspective, we need a distinction:

Survey methodology – the collection of practices for (official) statistics production

vis-à-vis

Survey (statistics) theory - a mathematical field, rooted in a central idea of statistical science:

From a part, make inference to the whole

Survey theory

- is mathematical
- the best of it has (over the years) had tremendous impact on practice
- taught only in few universities

Illustration: IASS jubilee commemorative volume 2001 (Landmark papers in Survey Statistics):

19 papers, almost all mathematical

Survey theory

A division within Survey Theory is:

- Design-based (probability sampling) theory, from 1930's
- Model–based theory, from 1970's, as in Small area estimation.

Survey theory

Classical (design-based) writers & pioneers :
W.G. Cochran, W.E. Deming,
M. H. Hansen,

They were (applied) mathematicians with a keen understanding of the practical exigencies of surveys.

Survey theory

Cochran, Deming, Hansen : Pioneers

"A ***theoretical statistician*** is one who guides his practice with theory. The theoretical statistician is the practical man, as he has a better guide for practice than the errors of his forefathers. Statistical theory shows how mathematics, judgement and substantive knowledge work together." (Deming, 1960)

Survey theory

has come a long way since 1950's
Is it today ***a mature science*** ?

Imre Lakatos (as cited by L. Laudan) :
A science reaches ***maturity***
when scientists in that field consistently ignore both anomalous problems and outside intellectual and social influences
and focus almost entirely on the mathematical articulations of research programmes

So then what is
Survey methodology ?

Cochran, Deming, Hansen :
Look through their classical books from around 1960 !

Survey methodology: The term is not there !
Imputation, small area estimation, editing: Also not there!
Nonresponse : Barely mentioned

What is
Survey methodology ?

Had you asked Cochran or Deming or Hansen around 1955, they would not have been familiar with the term "survey methodology".

Survey methodology is a "post-modern term" , necessitated largely by need to handle administratively the many problems arising in modern computerized, large scale data collection from increasingly un-cooperative human populations

In the classical era, 1940's to 60's,

Survey theory did exist.

Survey methodology did not exist
- as a term

Survey methodology

Today, in the 2010's ,
Survey methodology is:
"A collection of practices," each piece lending a certain support to one of the steps in statistics production process ("the statistical value chain")

- Is nevertheless extremely valuable
- Is systematically taught in very few places
  JPSM (USA) is a model;
  Europe lags behind

Survey methodology

Composed of great variety of courses
(e.g., at JPSM)

Data collection modes, Response behaviour, Interviewing, Pre-testing, Concern for data provider, Response burden, Confidentiality, and so on

With more mathematical orientation :
Imputation, Nonresponse weighting, Small area estimation, Editing, and so on

Survey methodology

The scientific underpinnings for ***survey methodology*** stem not only from ***statistical science*** ;
derive important elements also from :
(Cognitive) Psychology
Sociology (of interaction, of intergroup relations)
Economics
Political science

and not in the least,
Computer science

Survey theory vs. Survey methodology

To summarize:
To ***survey theory***
(as begun with Cochran, Deming, Hansen)
has in modern times become attached a balloon of practices and techniques,
necessitated by the complexity of modern times;

This has given us modern ***survey methodology***

The teaching of those fields :

Statistical science taught in many universities
Survey theory taught in very few universities
Survey methodology and official statistics production taught systematically in very few places,
but practiced in many

The statistician's responsibility

Can a statistician deliver ?

is the title of an article in

J. Official Statistics vol. **17** (2001), pp. 1 – 127

with 16 discussions

and a rejoinder by the authors,

R. Platek and C.E. Särndal

Can a statistician fulfill his/her promise (to society)?

It is to deliver reliable statistics - isn't it ?

Can a statistician deliver ?

The 16 discussants :

- Some say : Of course we cannot have a perfect theory for statistics production; the process is much too complex;
  they admit, if only reluctantly, that there is no objective measurement of accuracy in official statistics
- Others say : "the glass is more than half full"

Can a statistician deliver ?

Self-criticism:

We,Platek and I,
emphasized (perhaps too much)
the ***statistical science*** view,
its "idealistic obsession"
with "valid inferences to the population"

We did not point out that ***probability***,
the cornerstone of ***statistical science***,
is too limited a basis for ***official statistics***

Can a statistician deliver ?

Official statistics production
has "outgrown"
statistical science

Probability, the basis of ***statistical science***
is "too narrow" an instrument for
official statistics

Can we deliver ?

Probability and *probable error* play little or no role when people look at "published numbers" ; they see them as "the truth"

Franchet and Nanopoulos (1997), in : Statistical science and the European statistical system

"Very often, almost always, statistical results are presented as the pure truth, expressed through exact figures .. No confidence intervals are given, no methods of estimation are presented and no tests of significance are operated"

5. The future ?

Back to my "questions arising" :

- Could it be that official statistics production is so elementary, so lacking in established theory that questionable common sense can take over ?
- Information of unknown or doubtful accuracy produced against better knowledge, is it nevertheless *preferable* to high quality information, conforming to highest of standards?

5. The future ?

The NSI needs a protective armour, a shield for its mission to "produce official statistics" for the nation

In the past, this was not so necessary - the NSI was the unchallenged supreme instance of statistical competence - there was ***trust***

Today, the NSI is vulnerable.

5. The future ?

The NSI needs a shield for its ways of doing

Why ? Because there is

- competition, from sometimes less trustworthy competitors
- pressures from "high places"
- demands for more and more data on more and more things
- scarcity of resources

In the face of all this,
the nation's statistical high authority (the NSI) must demonstrate firm, competent delivery

5. The future ?

Today, the NSI refers to :
"A bundle of techniques"
in "the statistical value chain"
with "some theory here and there"
(from statistical and other sciences)

It is a weak protection.
It is too easy to poke holes in that defence,
by anyone who so chooses,
e.g., the government

5. The future ?

Information of unknown or doubtful accuracy, perhaps produced against better knowledge, is that nevertheless *preferable* (for society) to quality information, conforming to the highest of standards?

Not many have the opportunity (or the courage) to ask that difficult question

It lies at the heart of the Canada Census Incident.

5. The future ?

My *hope* for the future: That we be better able to show that sound, unifying, comprehensive theory can be brought in support of
"accurate and useful information"
for policy decisions
in the nation's interest

A *danger* lies in a more or less uncontrolled growth, an expanding balloon of "a collection of practices" , a fuzzy constellation without sharp contours (that is, more and more fragmentation)

5. The future ?

In particular, what can survey theory (the mathematically oriented survey science) contribute ?

References

Bohm, D. and Peat, F.D. (2000). Science, Order, and Creativity, 2nd edn. London: Routledge.

Franchet, Y. and Nanopoulos, P. (1997). Statistical science and the European statistical system: Expectations and perspectives. In Proc. Conference in honour of S. Franscini. Basel: Birkhäuser.

Groves, R. (1987). Survey research is a methodology without a unifying theory. Public Opinion Quarterly, 51, 156-172.

Lakatos, I. (1970). A chapter in: Criticism and the Growth of Knowledge. Cambridge Univ. Press.

Laudan, L. (1977). Progress and its Problems. Toward a theory of scientific growth. LA: Univ. of California Press.

Platek, R. and Särndal, C.E. (2001). Can a statistician deliver? J. Official Statistics, 17, 1 – 127 (with 16 discussions)

Assessing Disclosure Risk in Sample Microdata Under Misclassification

Natalie Shlomo[1]

Disclosure limitation methods for protecting the confidentiality of respondents in survey microdata often use perturbative techniques which introduce measurement error into the categorical identifying variables. Moreover, the data itself will often have measurement errors commonly arising from survey processes. Perturbation techniques are presented which allow for valid statistical inference when statistical agencies do not release the perturbation parameters. In addition, there is a need for valid and practical ways to assess the risk of identification through disclosure risk measures which take into account the misclassification. The probabilistic framework of the Poisson log-linear models for assessing disclosure risk is expanded to take into account the misclassification. This method will be shown in the context of probabilistic record linkage which is often used by statistical agencies to assess disclosure risk in perturbed datasets.

This talk is based on joint work with Prof. Chris Skinner from the Southampton Statistical Sciences Research Institute.

[1] *Southampton Statistical Sciences Research Institute, University of Southampton; E-mail:N.Shlomo@soton.ac.uk,*

PURSUING OPTIMAL COMPARABILITY IN A CROSS-NATIONAL SURVEY

Ineke Stoop[1]

Ineke Stoop is head of the Department of Data Services and IT, The Netherlands Institute for Social Research/SCP. She obtained her Ph.D. at Utrecht University for a thesis on survey nonresponse. She is a member of the European Statistical Advisory Committee (ESAC) and of the Central Coordinating Team of the European Social Survey. Ms Stoop is co-founder of the Dutch Platform for Survey Research and Laureate of the 2005 Descartes Prize for Excellence in Scientific Collaborative Research. Het main research interests are comparative social surveys and nonresponse. She has taught courses on comparative surveys and nonresponse as part of the ECPR summer school, as ESS training courses, and at Dutch universities, and recently co-authored a book on nonresponse in the European Social Survey.

Abstract

When the European Social Survey (ESS) was designed in the last decade of the previous century high quality and optimal comparability were two major goals. High quality entails all facets of total survey quality, including strictly random sampling and high response rates but also detailed translation procedures, multi-trait multi-method experiments, complete documentation and free and immediate access for all to the survey data. Optimal comparability was pursued by prescribing identical survey procedures in each country, e.g. a single mode of data collection, identical answer scales and coding instructions, and similar fieldwork requirements. In practice high national quality may sometimes conflict with optimal comparability, and reality interferes with the development and implementation of identical instruments: concepts differ across countries, different sampling frames are available, interviewers are free-lancers or employees and survey climates differ substantially. The presentation will discuss how optimal comparability is aimed for in the ESS.

[1] The Netherlands Institute for Social Research/SCP

CONSISTENT DOMAIN ESTIMATION IN MULTISURVEY SITUATION

Imbi Traat
University of Tartu, Estonia
e-mail: imbi.traat@ut.ee

It happens frequently nowadays that several surveys are carried out on the same population. Typically, some variables are common in two or more surveys. It is natural to require that estimates based on the common variables are consistent with each other in the different surveys. More specifically, in our case, population totals, or totals of larger domains, for certain variables are estimated in one survey, called the reference survey (RFS). Estimates based on the same variables but at a more detailed domain level are obtained from a second (later) survey, called the present survey (PRS). Consistency is required for the PRS estimates; domain totals for common study variables have to sum up to the corresponding estimated totals in the RFS. In a special case, the latter totals need no estimation; they are known exactly from registers. The RFS, or the registers, cannot by themselves provide the desired domain estimates due to the lacking domain identifiers.

We assume that the finite population U is divided into disjoint and exhaustive domains U_d, $d = 1, 2, \ldots, D$. The probability sample from U is s and its part falling into U_d is s_d. The aim is to estimate the vector of domain totals $\mathbf{Y}_d = \sum_{U_d} \mathbf{y}_k$, for $d \in \{1, 2, \ldots D\}$, consistently with $\mathbf{Y}_0$, the population total of $\mathbf{y}_k$, known exactly or estimated from another source. This means that we construct the weights w_{ck} such that the estimators

$$\hat{\mathbf{Y}}_d = \sum_{s_d} w_{ck}\mathbf{y}_k \text{ satisfy } \sum_{d=1}^{D} \hat{\mathbf{Y}}_d = \mathbf{Y}_0.$$

We use the calibration framework in a more general setting. Correspondingly, the expression for weights is

$$w_{ck} = w_k[1 + (\mathbf{Z} - \hat{\mathbf{Z}})' \mathbf{M}^{-1}\mathbf{z}_k] \text{ with } \mathbf{M} = \sum_s w_k \mathbf{z}_k \mathbf{z}_k', \tag{1}$$

where $\mathbf{Z} = \sum_U \mathbf{z}_k$, $\hat{\mathbf{Z}} = \sum_s w_k \mathbf{z}_k$ and $\mathbf{z}_k$ is a vector-variable. Our aim is achieved by choosing $\mathbf{z}_k$ and w_k in a suitable way. Three sets of weights are considered and compared in the presentation. An ordinary auxiliary variable calibration weight is a special case of (1), but it does not give the desired consistency with $\mathbf{Y}_0$.

References

[1] Särndal, C.-E., Traat, I. (2011). Domain estimators calibrated on information from other surveys. Submitted.

[2] Traat, I. (2010). Comparison of Calibration-based consistent domain estimators. In Carlson, Nyquist and Villani (eds), *Official Statistics - Methodology and Applications in Honour of Daniel Thorburn*, pp. 137-147. Available at officialstatistics.wordpress.com.

SAMPLING FOR HOUSEHOLD FINANCE AND CONSUMPTION SURVEY IN ESTONIA

Julia Aru[1]

Household Finance and Consumption Survey collects household-level data on household finances and consumption. It is the survey coordinated by European Central Bank and the first wave was conducted in several european countries in 2009-2010. The survey focuses on financial behaviour of households - real and financial assets, liabilities, income and employment, future pension entitlements, risk attitudes - but also consumption and saving. Due to the topics of interest of the survey, oversampling of the wealthy is desired in sample design.

In Estonia, the HFCS will be conducted in 2013, together with the second wave of HFCS in other participating countries. Planning of the survey is in progress and the presentation will discuss possible sample design for the survey, allowing for oversampling of the wealthy.

Considered design is two-phase startified sampling where first phase strata are geographical and formed by average income in the region and second phase strata are formed by personal income of those selected at the first stage. Income information for both phases comes from the Tax Board database.

Amount of oversampling achived with proposed design as well as problems and possible simplification will be discussed in presentation.

[1] Statistics Estonia

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR UNEQUAL PROBABILITY SAMPLING

Yves G. Berger[1] & Omar De La Riva Torres

Empirical likelihood (EL) (Owen, 1988) has first been introduced by Hartley and Rao (1968) under the name scale load approach. Since Chen and Qin (1993) suggested its first application in survey sampling, there have been many recent developments of empirical likelihood (EL) based methods (e.g. Rao & Wu 2009) and adaptive sampling (Salehi, *et al*. 2008). Chen and Sitter (1999) proposed a pseudo EL approach which can be used to construct confidence intervals for the Hájek (1971) ratio estimator (Wu & Rao, 2006).

Standard confidence intervals based upon a normal distribution can perform poorly when the sampling distribution is not normal. On the other hand, EL confidence intervals may be better in this situation, as EL confidence intervals are determined by the distribution of the data (Rao & Wu 2009). The range of the parameter space is also preserved. This may not be the case for standard confidence intervals based upon a normal distribution, as standard confidence intervals can have negative lower bounds for a positive point estimator.

The pseudo EL approach is not entirely appealing from a theoretical point of view (Rao & Wu 2009) as it is not a genuine EL approach, and it is not applicable to the Horvitz-Thompson (1952) estimator. We propose a true EL approach for unequal probability sampling without replacement based on Kim (2009) EL approach. We show that the profile EL function has asymptotically a chi-square distribution with one degree of freedom under a set of regularity conditions. Hence, the EL approach proposed can be implemented to construct confidence intervals for the Horvitz-Thompson estimator. We will also support our findings by a simulation study.

References

Chen J., Qin J. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. Biometrika 80:107–116 (1993)

Chen J., Sitter R.R. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex survey. Stat Sin 9:385–406 (1999)

[1]University of Southampton, United Kingdom

E-mail: Y.G.Berger@soton.ac.uk

Hájek, J.. Comment on “An essay on the logical foundations of survey sampling” by Basu, D. in Foundations of Statistical Inference (Godambe, V.P. and Sprott, D.A. eds.). p. 236. Holt, Rinehart and Winston. (1971)

Hartley H.O., Rao J.N.K. A new estimation theory for sample surveys. Biometrika 55:547–557 (1968)

Horvitz, D.G. And Thompson, D.J. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. 47, 663-685 (1952)

Kim, J.K. Calibration estimation using empirical likelihood in survey sampling. Statistica Sinica 19, 145-157 (2009)

Owen A.B. Empirical likelihood confidence intervals for a single functional. Biometrika 75:237–249 (1988)

Rao, J.N.K. & Wu, C. Empirical Likelihood Methods. In Handbook of Statistics: Design, Method and Applications: D. Pfeffermann and C.R. Rao. (editors). Elsevier (2009)

Salehi, M., Mohammadi, Rao, J.N.K, Berger, Y.G Empirical likelihood confidence intervals for adaptive cluster sampling, Environmental and Ecological Statistics. 17, (1), 111-123 (2008)

Wu C., Rao J.N.K. Pseudo empirical likelihood ratio confidence intervals for complex surveys. Can J Stat 34:359–375 (2006)

ALCOHOL CONSUMPTION SURVEY IN BELARUS

Bobrova Anastacia[1]

In Belarus, the population estimation of the alcohol costs is carried out on a regular basis as part of the household sample survey. Data available since 1995 – date of the first survey. Data on the average alcohol spending can be obtained as a percentage of the total household expenditures, as well as in monetary terms. So, according to the period 2005-2010. the share of spending on alcohol was about 2%. that was slightly lower the share of spending on health and culture, sport, and even higher education costs.

Nevertheless, the sample survey information should be treated very carefully. First, respondents were underestimating the consumption of alcohol significantly, particularly in sample surveys of European countries - more than 50%. Secondly, according to the national retail sale statistics, alcohol costs on average per household also increased by several times.

Problem of a representative sample of the analysis of alcohol consumption can be solved by increasing the sample size by 3 times or until the 15000-20000 at least once per 2 years. According to the households in the region figured out the coefficient of variation in alcohol consumption, which amounted to ungrouped data, more than 500%. Sample error is huge. Stratified sampling on the composition of households, groups of population (by sex and age, with the release of the youth and children) is required. A possible model of sampling: multistage stratified probability. A sampling unit is a private household.

[1]Institute of Economy of National Academy of Sciences, Minsk, Belarus

THE HOUSEHOLD SAMPLING IN OFFICIAL STATISTICS IN BELARUS

Natalia Bokun[1]

The main principles, characteristics and problems of two sample surveys of households (HH), conducted by state statistics in Belarus were considered: 1) expenses and income, 2) private subsidiary plots in rural areas. The sample unit is a household. Sampling design - territorial probabilistic multistage sampling. Nevertheless, the observations vary according to purpose, frequency, scope, methodology selection, methods of weighing.

The quarterly population survey, carried out since 1995, characterized by a fairly proven mechanisms of the development of sampling design, instrumentation and data processing. The first results of a survey of smallholders were obtained in 2010 and indicated the appearance of significant organizational and methodological problems: a high load of interviewers, the need for localization of the sample, the presence of atypical units, using a combination of methods to extrapolate each indicator questionnaire inadequate in some cases, background information household accounting.

Nowadays, the National Statistical Committee of the Republic of Belarus with foreign and national experts is the preparatory work on implementation of the Labour Force Survey (LFS). The sampling frame is 0.6% of households; the sampling design is territorial three-stage sampling. Using of individual weights (by sex, age, place of residence) is planned. The main difficulty in designing a methodological and software sample forming associated with the use of different weighting scheme, estimation of structural parameters of employment and unemployment.

[1] Belorussian State economic university (BSEU), Minsk, Belarus
e-mail: Nataliabokun@rambler.ru

TEACHING SURVEY SAMPLING THEORY AND METHODOLOGY AT THE OLES' HONCHAR DNIEPROPETROVSK NATIONAL UNIVERSITY

Iana Bondarenko[1]

The teaching program for the course on Survey Sampling Theory and Methodology at the Dniepropetrovsk National University is presented. It is intended for master students specialized in Applied and Theoretical Statistics. This course consists of 36 hours of lectures during the autumn semester in the fifth year of studies. The lectures on Survey Sampling are based mostly on the books by W. G. Cochran (1977), S. L. Lohr (1999), C.E. Särndal (1992), which are available in the electronic library of the Department of Statistics and Probability Theory of the University. The course covers the following topics:

1. Introduction to survey sampling
2. Simple random sampling without replacement
3. Bernoulli sampling
4. Systematic sampling
5. Sampling with replacement
6. Unequal probability sampling
7. Stratified sampling
8. Cluster sampling
9. Two-stage and multistage sampling
10. Estimation of functions of totals
11. Using auxiliary information

Students can study survey sampling by literature in Ukrainian. For example, *Lectures on Theory and Methods of Survey Sampling* of O. Vasylyk (2010), *Survey Sampling Technique* of O. Chernyak (2001), *Survey Sampling Methods* of V. Parkhomenko (2001) are the basic educational literature on the native language for the students.

References

Chernyak, O. (2001) *Survey Sampling Technique*. Kyiv (in Ukrainian).

Cochran, W. G. (1977) *Sampling techniques*. Wiley and Sons.

Lohr, S. (1999) *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove.

Parkhomenko, V. (2001) *Survey Sampling Methods*. Kyiv (in Ukrainian).

Särndal, C.-E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Schwarz C. J. (1997) StatVillage: An On-Line, WWW-Accessible, Hypothetical City Based on Real Data for Use in an Introductory Class in Survey Sampling. *Journal of Statistics Education*, 5(2). http://www/amstat.org/publications/jse/v5n2/schwarz.html

Vasylyk O., Yakovenko T. (2010) *Lectures on Theory and Methods of Survey Sampling*. Kiev (in Ukrainian)

[1] Oles' Honchar Dniepropetrovsk National University, Ukraine
e-mail: yanabondarenko@ua.fm

OUTLIER DETECTION IN BUSINESS SURVEYS

Baiba Buceniece[1]

One of the first steps in processing survey data is detection of outlaying observations (outliers). Outliers may lead to model misspecification, biased parameter estimation and incorrect results. It is therefore important to identify them prior to modeling and analysis.

This paper describes the outlier detection method in business sample surveys in Central Statistical Bureau of Latvia. Currently is used an automated method, which is proposed by M.Last and A.Kandel[1]. This method is based on fuzzy set theory.

To automate the outlier detection process, Last and Kandel represented the *comformity* of an attribute value by the membership function, which agrees with the definition of *fuzzy measure*. Given a set of distinct values, a value is considered *conforming from below* (*conforming from above*), if it is close enough to the succeeding (preceding) values respectively. Each conformity measure depends on the so-called "distance to density" ratio. For the conformity from below, this is the ratio between the distance from each value to the subsequent value and the average density of subsequent values.

References

[1] Mark Last, Abraham Kandel, *Automated Detection of Outliers in Real-World Data*, 2001.

[1]Central Statistical Bureau of Latvia

ON SOME ASPECTS OF THE STUDY COURSE "SURVEY SAMPLING"

Natalja Budkina[1]

Sample surveys are essential tools in a modern society to provide accurate information to politicians, businesses and the general public about living conditions and opinions. The course "Survey Sampling" is given in the programme of professional studies in Mathematical Statistics at the University of Latvia. This programme has been running at the Department of Mathematics, University of Latvia since 1997/1998 academic year. Now it is 4,5 years long and it is valued at 180 national credit points (70 of them are common with Bachelor Programme in Mathematics).

The special course in Survey Sampling was introduced at the Department of Mathematics in 1996 by Dr. Math. J. Lapiņš. It was intended for the students of the programme of the professional studies in Mathematical Statistics in the fourth-fifth year of studies. For the reaccreditation of this programme in 2007 the course was reworked. The changes were not very essential, but they were important taking into account the development of Survey Sampling Theory and Methodology during the last years. Now this course consists of 54 hours of lectures and 10 hours of practical lessons during the autumn semester in the fourth year of studies. The talk presents a short description of this course, its theoretical part and practical works.

During the whole period of teaching the course Survey Sampling it has been popular among the students. Every year there is at least one student choosing a theme on Survey Sampling for his/her probation work or Diploma paper.

It is quite a problem to organize the probation work for students on a high level. The tasks which the students have to fulfil are essentially different but a lot of students choose themes connected with survey sampling. A long period (26 weeks) of probation work gives the possibility for the students to participate in real sampling projecting, carrying it out and its analysis. Some examples of the probation works are presented in the presentation.

Diploma work constitutes an essential part of the study programme. Since the course "Survey Sampling" is intended for the students of the programme of professional studies, most of the Diploma Papers on this theme have practical purpose. Since 1996 25 Diploma Theses on Survey Sampling have been defended by students of the speciality "Mathematician-Statistician" of the University of Latvia. The problems concerning the Diploma Papers are mentioned in the talk.

The perspectives of development of the course "Survey Sampling" are considered in the presentation.

[1] University of Latvia, Riga Technical University, Latvia

NONRESPONSE BIAS IN THE *SURVEY OF YOUTH PERCEPTION OF SCIENCE AND TECHNOLOGY IN BOGOTA*

Edgar Mauricio Bueno Castellanos*

Resumen

The Colombian Observatory of Science and Technology —OCyT— developed, in 2009, a survey about the perception of Science and Technology in students of the last two years of high school in Bogotá, Colombia. The survey sampling design was stratified according to the nature of school (official or private). During the data collection stage, two main sources of nonresponse were detected. The first one, as a consequence of the important difference in the response probability according to the nature of school: the survey was implemented in the 16 official schools included in the original sample (100 %), while only 13 out of 31 private schools (42 %) allowed to collect information. The second source corresponds to students who belong to schools in which access was allowed, but did not assist during the days when survey was applied. Estimates, initially, were obtained modifying the original sample sizes by those observed. Subsequently, it was decided to obtain new estimates taking into account the nonresponse effect; to achieve this goal, the values corresponding to item nonresponse were imputed using the methodology of the *nearest neighbor* and the calibration method was used for unit nonresponse. The results obtained for both cases don't show visible differences, especially when estimating a ratio; even though, some great differences were observed when estimating totals. Although it is impossible to determine with certainty which methodology was more precise, results obtained using the combination imputation-calibration are more reliable because it considers auxiliary information, which is not present in the first estimation methodology.

Keywords: *Sampling design; nonresponse bias; calibration.*

*Colombian Observatory of Science and Technology -OCyT-

PERCEPTION ABOUT CHARACTERISTICS OF SCIENTISTS, RESULTS OF THE *SURVEY OF YOUTH UNDERSTANDING OF SCIENCE AND TECHNOLOGY IN BOGOTÁ*

Author: Edgar Mauricio Bueno Castellanos. Statistician.

Institution: Colombian Observatory of Science and Technology -OCyT-.

The Colombian Observatory of Science and Technology -OCyT- developed, in 2009, a survey about the perception of Science and Technology in students of the last two years of high school in Bogotá, Colombia. A set of questions inquired about the perception of these students of the characteristics of a scientist. Answers were analyzed using multivariate data analysis tools. Three classes of students were identified: students that think of a scientist as a person with special intellectual capabilities, students that see scientists as isolated people and, a third class, composed by students that think of scientist as ordinary people. Given that the sampling design -stratified according to the nature of school (official or private)- should not be ignored at the estimation stage, the analysis was carried out taking into account the selection probabilities assigned to schools. In addition, because of the high nonresponse, especially in private schools -data collection was allowed in 13 out of 31 private schools (42 % of schools, 53 % of expected students)-, it was decided to impute values corresponding to item nonresponse using the methodology of the *nearest neighbor* and use calibration methods for unit nonresponse, reducing the nonresponse bias.

Keywords: *Science and technology perception; multivariate analysis; sampling design.*

ON AN OPTIMAL BOUND FOR THE VARIANCE OF SAMPLE MAXIMUM

Andrius Čiginas[1]

Let $\mathcal{X} = \{x_1, \dots, x_N\}$ denote measurements of the study variable x of the population $\{1, \dots, N\}$. Let $\mathbb{X} = \{X_1, \dots, X_n\}$ denote measurements of units of the simple random sample of size $n < N$ drawn without replacement from the population. Let $X_{1:n} \le \dots \le X_{n:n}$ be the order statistics of $\mathbb{X}$. We are interesting in an optimal bound of the form $\mathbf{Var}\, X_{n:n} \le a_{n;N}\, \mathbf{Var}\, X_1$, where $a_{n;N}$ may depend on n and N only. Here optimality means that there exists a nontrivial population where equality is attained.

Let us mention few known results, which we extend to samples drawn without replacement. In the case of independent identically distributed (i.i.d.) observations Papadatos (1995) showed that $a_{n;N} = n$. The same constant appears in the case of arbitrarily dependent identically distributed observations, see Rychlik (2008). For i.i.d. samples, additionally assuming that X_1 has a symmetric distribution, Moriguti (1951) obtained $a_{n;N} = n/2$.

We note that for samples drawn without replacement the optimality constant $a_{n;N}$ is the same for the sample minimum. Similar bounds for the variance of other order statistics will be also discussed at the conference.

References

Moriguti, S.: Extremal properties of extreme value distributions. Ann. Math. Statist. 22, 523–536 (1951)

Papadatos, N.: Maximum variance of order statistics. Ann. Inst. Statist. Math. 47, 185–193 (1995)

Rychlik, T.: Extreme variances of order statistics in dependent samples. Stat. Probab. Lett. 78, 1577–1582 (2008)

[1]Vilnius University, Lithuania

COMBINING DATA FROM LABOUR FORCE SURVEY AND REGISTER OF UNEMPLOYMENT SURVEY

Andris Fisenko[1]

The main goal of the research is to compare data of Labour Force Survey (LFS) 2010 with Unemployed register. Using Unemployed register as auxiliary information I will link with LFS data and try find differences in personal level. Data will be linking using unique person identity number. From LFS data will be used questions according person unemployed status. While information from Unemployed register will be used about corresponding period. Since this information are used as calibration for data weighting its important to know real situation. Its possible to find different errors, mistakes and other aspects of misleading information comparing both data sets. Its possible because in LFS households (persons) are interviewed 4 times. Results will show as the data accuracy.

Depending of results one of options can be minimized the burden of interviewers and decrease numbers of questions about unemployment. Second options shows as what we can use from Unemployed register instead to asking respondents.

[1]Central Statistical Bureau of Latvia

COMPOSITIONAL ESTIMATION: THE CASE OF NUMEROUS INFORMATION SOURCES

Oleksandr Gladun [1]

Classical composition estimation formula has two elements: direct estimation and indirect estimation. Each of them has its own weighting coefficient.

But there may be situations, when there are more then two estimates. In this case the problem of weighting coefficient determination for each estimate arises.

There is proposed the general compositional estimation model for the case of unlimited quantity of small area estimations, as well as weighting coefficient calculation formula. This provides a possibility to use any number of direct and indirect estimations together.

Also there are considered model application conditions: data harmonization of all surveys and belonging of direct sample estimations to the same total population.

The model usage is checked on the data of real sample surveys.

[1] Ptukha Institute for Demography and Social Researches, Ukraine

Spatially balanced sampling through the Pivotal method

Anton Grafström[1], Niklas L.P. Lundström[2], Lina Schelin[2]

A simple method to select a spatially balanced sample using equal or unequal inclusion probabilities is presented. For populations with spatial trends in the variables of interest, the estimation can be much improved by selecting samples that are well spread over the population. We show how the Pivotal method, introduced by Deville & Tillé (1998), can be used to select samples with a high degree of spatial balance. We incorporate the spatial aspect by introducing distance between units in the design. Most spatial applications naturally concern populations spread in one, two or three dimensions. However, the method can be used for any number of dimensions since all that is needed is a measure of distance between units. The main idéa is to create a strong negative correlation between the inclusion indicators of units that are close in distance. In that way units that are close in distance seldom appear together in a sample, which creates a well spread sample. Analysis and examples indicate that the suggested method achieves a high degree of spatial balance and is therefore efficient for populations with trends.

References: Deville, J-C. & Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika* **85**, 89-101.

[1]Department of Forest Resource Management, Swedish University of Agricultural Sciences, SE-90183 Umeå, Sweden

[2]Department of Mathematics and Mathematical Statistics, Umeå University, SE-90187 Umeå, Sweden

QUALITY ASSESSMENT FOR INVESTMENT SURVEY

Oksana Honchar[1]

The development of the information society puts in the forefront new demands of statistical information and particularly its quality. The problem of quality assurance of statistics cannot be solved by isolated unsystematic approaches. Instead, a methodical approach to quality management is necessary to apply.

Ukrainian statisticians have understood the importance of quality management system development and its implementation. At present many projects in Ukraine within this specific sphere are established and many others are scheduled in the future. More sophisticated statistical methods are implemented due to the growing use of sample surveys. So, elaboration of quality assessment methodology for sample surveys is very actual problem for ukrainian statistics.

Statistical surveys of capital investments are conducted on a quarterly as well as annual basis. In these surveys only the most representative enterprises have so far been observed so quality control of the obtained estimates have been very problematic. Currently methodological support of the capital investment sample survey is being elaborated. Therefore the aim of the report is to evaluate the quality of the planning process for this survey.

References

Eurostat: Handbook on Data Quality Assessment Methods and Tools. Wiesbaden (2007)

Eurostat: Handbook on Improving Quality by Analysis of Process Variables

Eurostat: Quality Improvements of the Survey Processes. Piraeus (2009)

[1] National Academy of Statistics, Accounting and Audit, Scientific and Technical Complex for Statistical Research, Ukraine

SOME ASPECTS OF DATA QUALITY ASSESSMENT OF BUSINESS TENDENCY SURVEYS ON UKRAINIAN CONSTRUCTION COMPANIES

Liudmyla Iashchenko[1]

Since 1997 Business Tendency Surveys on construction companies have been conducted in the Research Institute of Statistics (later - Scientific and Technical Complex of Statistical Research) together with the State Statistics Committee of Ukraine on a regular quarterly basis, along with surveys such economic sectors as industry, trade, transport and agriculture. Quarterly more than 300 construction companies are surveyed. Conducting of the surveys provides an opportunity to obtain operative and accessible information to different specialists that is ahead of other statistical data on 1 or 2 quarters.

To improve the data quality assessment of the results of Business Tendency Surveys on construction companies in the STC of Statistical Research not only the variety of errors are taken into account, but also indicators that are used in sociology and psychology, such as the coefficient of reproducibility of the scale and Gilford's φ-coefficient to estimate the validity. Such algorithm for estimating quality of the results of Business Tendency Surveys is used: 1) analytical estimation of the deviations between the results of surveys and relevant statistics, 2) calculation of the number of missing responses, 3) logic control of responses; 4) estimation of the reliability of survey results; 5) assessment of the reproducibility of the scale; 6) checking the validity of the data. To estimate the relevance of the data additional survey of respondents is carried out.

Practical calculations obtained on the basis of Business Tendency Surveys on construction companies of Ukraine confirmed the improvement of the quality of the results.

[1] Scientific and Technical Complex of Statistical Research, Ukraine

ESTIMATION IN THE PRESENCE OF NONRESPONSE AND MEASUREMENT ERRORS

Maiki Ilves [1]

In the classical survey sampling theory the only errors considered in the estimation are sampling errors. However, often nonsampling errors are more influential to the properties of the estimator than sampling errors. This is recognized by the practitioners and researchers and a lot of literature regarding nonsampling errors has been published during last two decades, especially regarding nonresponse error. Most of this literature handles one kind of nonsampling error at a time, although in real surveys more than one nonsampling error is usually present.

In this paper two kinds of nonsampling errors are considered at the estimation stage: nonresponse and measurement error. A calibration estimator corrected to the bias due to the measurement errors is introduced. In official statistics calibration estimator is often used to reduce the nonresponse bias. In order to estimate the bias due to the measurement errors probability editing is carried out. The theoretical variance, as well as the variance estimator, of the given estimator is derived. The results of a small simulation study are also presented.

Keywords: nonsampling errors, calibration estimator, bias estimation

[1]Department of Statistics, Örebro University, Sweden

FAMILY OF ESTIMATORS OF POPULATION VARIANCE IN SUCCESSIVE SAMPLING

Nursel Koyuncu[1]

Successive sampling consists of selecting sample units on different occasions such that some units are common with samples selected on previous occasions. We consider sampling on two occasions from a finite population $U = \{1, \ldots N\}$ of size N. Associated with the *i*th unit are, the study variable, y_i, and auxiliary variable, x_i. Note that on the first occasion, the study variable y is called the auxiliary variable x. Using simple random sampling n units are selected on the first occasion. A random sub sample of $m = n\lambda$ units are retained for use on the second occasion, while a fresh simple random sample of $u = (n - m)$ units are drawn on the second occasion from the remaining $(N - n)$ units of the population. In this scheme, drawing sample at different time point gives more reliable estimates than one-stage sampling. We can take into account the change of the characteristics over different occasions.

In this paper we have intended to improve the precision of variance estimates at the current occasion. We have proposed a family of estimators of population variance and the expressions of bias and mean square error are derived in successive sampling. To analyze its properties we have carried out an empirical study based on real populations.

References

Koyuncu N. Kadilar, C., (2010). On Improvement in Estimating Population Mean in Stratified Random Sampling, Journal of Applied Statistics, 37, 6, 999-1013.

Singh, H.P., Tailor R., Singh S., Kim J.M., (2011). Estimation of Population Variance in Successive Sampling, Qual Quant, 45, 477-494.

Singh, G.N., Priyanka K., (2010). Estimation of Population Mean at Current Occasion in Presence of Several Varying Auxiliary Variates in Two-Occasion Successive Sampling,Statistics in Transition, 11, 1, 105-126.

[1] Hacettepe University, Turkey

STRATEGY FOR ESTIMATION OF PARAMETERS IN HOUSEHOLD SURVEY

Danutė Krapavickaitė[1]

Because of lack of the registers to build a sampling frame the statistician has to use the data bases which are available and to create complex sampling designs and complex estimation methods.

One of the situations of such kind arises in a survey of households (or dwellings) when their list is not available, and population register of the individuals is used for a sampling frame. A total of a study variable defined for households (or dwellings) has to be estimated. At the same time there is requirement on the weights of the estimator to estimate the size of the population of individuals and some other totals without errors.

Sampling of households through individuals is often used in official statistics when household register is not available. Individuals are selected with equal probabilities with or without replacement, and their households are included into the sample, excluding the repetitions, until fixed number n of different households has been selected, and these households constitute a sample. The larger is the household, the higher its inclusion probability into the sample. The sampling design described above is exactly the sampling design which has been studied by Rosén [2] and called successive *pps* sampling. We use it as a first-phase sampling design. The second phase sampling design, stratified simple random sampling with sampling rate inverse to the household size, described in Traat and Ilves [3], is included into our strategy. It allows to equalize household inclusion probabilities and to reduce the variance of the estimator of the population total. The way of calibration for two phase sampling design to meet some restrictions for the estimation weights, described in Estevao and Särndal [1], is used at the estimation stage.

References

1. Estaevo, V, Särndal, C.-E. A New Face on Two-Phase Sampling with Calibration Estimators. *Survey methodology*, 2009, 35, 3-14.
2. Rosén B. *Asymptotic Theory for Order Sampling*. Statistics Sweden R&D Report 1995:1.
3. Traat, I., Ilves, M. The Hypergeometric Sampling Design, Theory and Practice. *Acta Appl. Math.*, 2007, 97, 311-321.

[1]Vilnius University Institute of Mathematics and Informatics, Vilnius Gediminas Technical University, Lithuania, Statistics Lithuania

A mixed mode design vs one-mode design

Seppo Laaksonen[1]

Enticement to use web surveys is becoming more common, since this mode is less expensive than the alternatives. On the other hand, no-one trusts in web as the only mode, since non-response is expected to be too high even though a good sampling frame can be created. Hence, a mixed mode strategy is proposed. There are different approaches to mixed mode design.

We test such an approach that the data collection will start via web and after a certain short period, CATI will make attempts to complete the fieldwork. Finally, we compare the successfulness of this mixed mode strategy against pure CATI. The reason is that this survey has been compiled using CATI but if a new mixed mode strategy is good, this will be used in future. So, we have to compare the successfulness of both approaches in a good way. This also requires to design a sample well. Our target population is the same as used in the regular survey that does not cover the full population but the people with telephone access only.

The paper explains our sampling design and presents the basic results from the test. The unit non-response of the comparable surveys (mixed vs one-mode) was not much different from each other but the questionnaires are not in all aspects comparable. Nevertheless, our results are promising and we propose to use a mixed mode in a regular survey too but some further developments are needed.

References

de Leeuw, E. D., (2005): To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* 21. 2, 233-255.

de Leeuw, E. D and Dillman, D.A, Hox, J.J. (2008): Mixed mode surveys: When and why. In: *International Handbook of Survey Methodology* (Eds. de Leeuw, E.D., Hox, J.J., Dillman, D.A). 299-316. Taylor & Francis Group, LLC. New York.

Jäckle, A., Roberts, C. and Lynn, P. (2009). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review* 78, 1, 3-20.

Revilla, M. (2010). Quality in Unimode and Mixed-Mode Designs: A Multitrait-Multimethod Approach. Survey Research Methods 4, 151-164. http://www.surveymethods.org

[1] University of Helsinki, Finland

INSTRUMENT VARIABLE SELECTION IN THE CALIBRATION ESTIMATOR UNDER NONRESPONSE

Thomas Laitila[1] and Carl-Erik Särndal[1]

The calibration estimator suggested by Särndal and Lundström (2005) uses sample and/or population level information on a set of auxiliary variables to adjust the design weights, in an effort to reduce nonresponse bias. The calibrated weights are based on known or unbiased estimates of auxiliary population totals. When attached to the study variable(s), observed only for the responding units, the effect of the calibrated weights is to reduce the bias, compared with less attractive alternatives, such as the simple expansion estimator. Several well known estimators, e.g., the post stratification and the generalized linear regression estimators, are special cases of the calibration estimator. Important for effective bias reduction is the strength of relationship existing on the one hand between the auxiliary variables and the study variable, on the other hand between the auxiliary variables and the response probabilities. Those aspects are reflected in methods for auxiliary variable selection developed in Särndal and Lundström (2005, 2008, 2010), Särndal (2011) and Schouten (2007).

Earlier contributions have focused on bias reduction using weights that are "standard" in the sense that the auxiliary vector is also the instrument vector used for computing the calibrated weights. In this paper, the objective is instead to reduce bias by choosing an instrument variable vector other than the auxiliary variable vector itself. Using an expression in Särndal and Lundström (2005) for the bias of the calibration estimator, we can state a condition on the instrument vector under which the calibration estimator is nearly unbiased. This condition allows us to establish a connection between the calibration estimator and the procedure proposed by Heckman (1979) for dealing with selection bias in regression analysis. The connection gives a procedure for formulating the instrument vector in the calibration estimator of finite population quantities. The remaining bias of the resulting calibration estimator is evaluated by an empirical example.

References

Heckman, J.J. (1979). Sample Selection as a Specification Error, *Econometrica* **47:1**, 153-161.

Särndal, C.E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York: Wiley.

Särndal, C.E. and Lundström, S. (2008). Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator. *Journal of Official Statistics*, **24**, 251-260.

Särndal, C.E. and Lundström, S. (2010). Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias. *Survey Methodology*, **36**, 131-144.

Särndal, C.E. (2011). Three factors to signal nonresponse bias; with applications to categorical auxiliary variables. To appear.

Schouten, B. (2007). A selection strategy for weighting variables under a not-missing-at-random assumption. *Journal of Official Statistics*, **23**, 51-68.

[1] Örebro University and Statistics Sweden

REPRODUCTIVE HEALTH SURVEY IN BELARUS: THE POSSIBILITY OF HOLDING

Anna Larchenko[1]

The existing statistical information on a population health state should form the base for co-ordination of the health protection policy, improvement of a population health state, for planning health care services and social security. In these purposes researches of the population's health are carried out in many countries. They are based on data from statistics, registers and population surveys, usually health interview surveys. Some information (for example, how the person estimates the health by himself, whether he is satisfied by public health services etc.) can be received only with the usage of survey sampling.

Reproductive health is the major component of the general health and by that takes the central place in development of each person. Being not only health reflection at children's and adolescents' age, it creates a basis for maintenance of health after the lapse of reproductive years of a life, both women, and men, and defines the consequences transferred from generation to generation.

In Belarus research of a condition of reproductive health represents the greatest interest as in the country there is a number of problems, such as: low birth rate, men's supermortality rate at reproductive age, high level of infertility, low male life expectancy etc. Thus, according to the official statistics, in 2009 total fertility rate was 1.44; the percentage of infertile couples – 16-17%; mortality of men in reproductive age was 5.24 per 1000 men of reproductive age (women – 1,78 ‰), male life expectancy – 64.7 years.

The author proposes holding a special sample survey of health of the population at reproductive age in Belarus. The preparatory stage of this survey can serve as a test of mini-survey in Minsk.

References

Bethlehem, J.: Applied Survey Methods: A Statistical Perspective, 43-122 (2009)

Reproductive Health Survey: Romania 2004. Sum. Rep., 14-16 (2005)

[1] Belarus State Economic University, Belarus

SIMULATION STUDY OF SAMPLING DESIGN IN LABOUR FORCE SURVEY

Mārtiņš Liberts[1]

A common task for a statistician in official statistics is planing and maintenance of the sampling design used for a sample survey. Usually there are two main goals set for a sampling design – high precision of survey estimates and low expected survey cost (variable cost in this case).

Precision of survey estimates and expected survey cost depend on a chosen sampling design. For example, clustering of sample units will decrease a precision but it will decrease survey cost at the same time. This is common for surveys where survey mode is personal interviews done by interviewers.

Sampling design used for Latvian Labour Force Survey (LFS) is a complex design [1]. It is not an easy task to evaluate precision and expected cost of a survey because of the complexity of a survey design. Simulation experiment is considered as a possible tool for evaluation of precision and expected cost.

Artificial population data are required for simulation experiment. The data should represent the target population of a survey as close as possible. The data from the Latvian Population Register and the data from LFS are used to derive artificial population data.

The setting of a simulation experiment can be described as repeatedly sampling from the artificial population by specified sampling design. Estimation of population parameters and survey cost is done using sample data in each iteration.

The results gained from different simulation experiments can be used to compare different sample designs regarding precision and expected cost. Such information can be used for evaluation of the current sampling design and as well for planing of changes in sampling design.

References

[1] Liberts M. *The Redesign of Latvian Labour Force Survey*, pages 193–203. Official Statistics – Methodology and Applications in Honour of Daniel Thorburn. Department of Statistics, Stockholm University, Stockholm, Sweden, 2010. Available at http://officialstatistics.wordpress.com/.

[1] Central Statistical Bureau of Latvia

CONSISTENT ESTIMATION OF CROSS-CLASSIFIED DOMAINS

Kaur Lumiste[1]

Domain estimation has become an important area in survey sampling, but a lot of problems associate with it. For example small sample size in domains produce inaccurate estimates. Another problem with domain estimation is the lack of consistency between different surveys. The results of one survey do not coincide with the results of another survey done earlier or simultaneously, although the same variable is under study. We overcome the inconsistency by applying two new methods - AC calibration or repeated weighting (RW).

One goal of this paper is to give a short overview of the AC calibrated estimator proposed by Traat and Särndal (2009) and of the RW method developed in Statistics Netherlands (see Kroese and Renssen (1999), Knottnerus and van Duin (2006)). Also, formulas for a specified case, for the cross-classified domains are given.

We assume that there are two sources of information of the study variables (either surveys or registers). But the problem is that one source has information on domains formed by certain categorical variable, not considered or not identified in the other source. Instead, this second source has information on domains formed by another categorical variable. We are however interested in domains that form from the cross-classification of the previous categorical variables. A survey is done regarding these new domains, but the domain estimates will probably be inconsistent with earlier information. So we take the earlier information and insert this as marginals to our 2-way table, and demand that the new found domain estimators are consistent with the marginals. To achieve this we apply AC-calibration or repeated weighting.

The formulas of AC calibration and RW for the cross-classified domains case are novel and are also tested in a simulation study. Simulations were done on an artificial population composed of real data from the Estonian Household Survey.

References

Knottnerus, P., van Duin, C. (2006) Variances in Repeated Weighting with an Application to the Dutch Labour Force Survey, *Journal of Official Statistics*, 22, 565-584.

Kroese, A.H., Renssen, R.H. (1999) Weighting and Imputation at Statistics Netherlands, *Proceedings of IASS Satellite Conference on Small Area Estimation*, Riga:Latvia, 109-120.

[1] Estonian-Swedish Mental Health and Suicidology Institute, Estonia

Traat, I., Särndal, C.-E. (2009) Domain estimators calibrated on information from other surveys, *Research Report* No. 2009-1, vol. 15, Department of Mathematics and Mathematical Statistics, Umeå University, Sweden.

WALD AND WILSON INTERVAL ESTIMATION OF CRIMINAL OFFENCES IN SMALL AREAS

Måns Magnusson[1]

Binary data are central in estimating crime and victimizations rates. When you want to produce interval estimates for small areas and when the crime types are quite rare, interval estimates based on approximations to the normal distribution cannot always be used.

Earlier research (Brown, Cai, DasGupta 2001) suggests that the classical Wald interval, based on normal approximation, doesn't work very well in the situation of rare events and small samples. The two-sided Wilson interval has been suggested as an alternative.

In this study the Wilson interval has been adapted for a generalized regression estimator and is examined through a design-based simulation study based on real life data from the Swedish Crime Survey. The Wilson interval and the Wald interval have been studied with regard to coverage percentage. Different crime types as well as two different models with different strength of the auxiliary variables have been studied.

One result is that the classical Wald interval performs badly whith regard to coverage, even for quite large sample sizes such as $101 < n < 349$ for the most rare offences as sexual assault and robbery. The Wilson interval, adapted for the GREG estimator, proved to work much better. For less rare offences such as "Crime against persons" the Wilson interval works for all domain sizes. For the less common crime types, such as robbery and sexual assault, the Wilson interval works well for sample sizes with $n > 100$.

[1] Stockholm University and Smittskyddsinstitutet

Application of the Panel Data models in Small Area Estimation

Vilma Nekrašaitė-Liegė[1]

Nowadays the official statistics repeats the same surveys from year to year, so for the most of the population elements it is possible to get information for the same variable in several time periods. It means that for many surveys panel-type data is available. Also, in some cases it is possible to use information collected from the other sources (tax offices, jobcenters and etc.). Such dataset of a large amount of auxiliary information might improve the quality of the estimation strategy (a pair comprising a sample design and an estimator) as compared with a strategy based on a current sample alone.

The use of panel-type data in estimation strategy means, that a prediction theory based on a superpopulation model is used. A superpopulation model can be used not only in estimation stage, but also in sample selection stage. Such use of the superpopulation model is discussed by Royall (1970), Nedyalkova, Tille (2008). They used linear regression model as a superpopuliation model.

In this research, a basic superpopulation model is a panel data model with random effects (Hsiao, Pesaran (2004)). The accuracy of estimation strategies is investigated for the small areas. The focus on small area is made because nowadays the demand for accurate statistics on domains (including small areas) of the population is growing. Also for the small area estimation an important aspect is to choose the right estimator and model (Lehtonen, Sarndal and Veijanen (2003, 2005)).

Several models and estimation strategies are compared by a computer experiment. The experiment is based on the real data (*Lithuanian survey on short-term statistics on service*) taken as a true population.

References

Hsiao, C., Pesaran, M.H.: Random coefficient panel data models, Cambridge Working Papers in Economics No. 0434, Fac-

[1]Vilnius Gediminas technical university, Statistics Lithuania, Vilnius, Lithuania

ulty of Economics, University of Cambridge, 2004, Web site: http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0434.pdf

Lehtonen, R., Sarndal, C.-E. and Veijanen, A.: The effect of model choice in estimation for domains, including small domains, *Survey Methodology* 29:33-44, 2003.

Lehtonen, R., Sarndal, C.-E. and Veijanen, A.: Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains, *Statistics in Transition* 7:649-673, 2005.

D. Nedyalkova, Y. Tille, Optimal sampling and estimation strategies under the linear model, *Biometrika* 95[3]:521-537, 2008.

R. M. Royall, On finite population sampling theory under certain linear regression models *Biometrika* 57[2]:377-387, 1970.

Real estates with fishing rights in Sweden - a survey design

Jens Olofsson
Department of Statistics
Örebro University
S-701 82 Örebro
Sweden
jens.olofsson@oru.se
Statistics Sweden S-701 89 Örebro
Sweden jens.olofsson@scb.se

April 25, 2011

Abstract

Fishing rights is a right of disposal referred to real estates. There exists no collective source of information on real estates with fishing rights or owners of such, so called fishing right owners, in Sweden. Hence, the knowledge is limited and more knowledge and insight is needed.

This paper presents a survey design used in order to find and estimate the total number of real estates with fishing rights as right of disposal in Sweden 2008 and the main results of the survey. The sampling design used is based on a disproportional stratified sampling design utilising the available auxiliary information in the sampling frame based on the Swedish national land register. The auxiliary information has also been used in the choice of sampling design within stratum as well as in the estimation stage.

Keywords Survey design, sampling design, probability proportional to size sampling, rare population, estimation

QUANTIFICATION THE FACTORS OF STATISTICAL WORK LABOR INPUT USING THE METHODS OF SAMPLE SURVEYS

Julia Orlova

The research work is connected with an estimation of labor input and definition cost of statistical works. Statistical work is intellectual therefore there is a problem of estimation of exit product cost. That estimation is necessary for definition of cost of the services rendered by the state statistics on a paid basis, and also for definition of volumes of budgetary financing. In the course of research methods of biographical inspection of workers of the state statistics and a selective photo of the working day have been used. By means of statistical methods the quantitative estimation of the factors influencing labor input of statistical works has been calculated. Also the specifications of expenditures of labor by the allocated kinds of statistical works are calculated.

As technological operations carried out by employees of statistics with the input and output information vary considerably, so it is advisable to quantify the factors influencing the complexity of the statistical work, with a preliminary subdivision of the statistical work to: 1. work with the input information; 2. work with the output information. Working with the input information includes the collection operation, registration, systematization of statistics. Working with the output information includes such transactions as processing and presentation of summary statistics on economic, demographic, social and environmental situation in the Republic of Belarus, their accumulation and storage.

According to a sample survey conducted in government statistics revealed 4 factors that influence the complexity of processing the input information:

1. The complexity of information.
2. The number of signs of indicators.
3. The average occupancy of each form, which also determines the time of supervision and consultation with businesses on issues that arise in the process of control.
4. The volume of information processed for the year determined for each type of input information as a product of the number of legal entities and their separate divisions, reporting to form, on its periodicity.

And the first three factors are qualitative (intensity) and affect the processing time per report (or other media input), whereas the fourth factor is a quantitative (extensive) and takes into account the amount of information to be processed.

According to results of the questionnaire the factor of the amount of output information was also calculated. This factor affects the labor input of statistical work related to the results of statistical information provision at a higher level.

References

Developing a national strategy for statistical development of the Republic of Belarus. Minsk, 5-6 October 2006

The main provisions of the program of socio-economic development of Belarus for 2006-2010 / Sovetskaya Belorussia from 24.02.2006 P. 9-16.

KERNEL IMPUTATION - A METHOD TO REDUCE BIAS OF HOT DECK IMPUTATION

Nicklas Pettersson[1]

Consider a sample from a finite population with missing values, where the goal is to estimate some finite population characteristic, typically a mean or a total. One way of handling the missing values is hot deck imputation. Each donee unit that has some missing values is then matched up with a pool of donor units, based on the similarity between values that are observed both on the donee and its potential donors. The missing values are then filled in by copies from corresponding observed values on units that are (randomly) drawn from the donor pool.

Hot deck imputation is good at preserving distributions among variables, and therefore provides robustness to nonlinear relationships. Estimates may however suffer from bias, if the continuity of the observed variables is not sufficiently accounted for in the matching of the donee to its potential donors, for example if continuous variables are categorized. The bias is especially evident if the donee is located at the boundary of the observed data.

By incorporating several ideas from kernel density estimation, we propose how to reduce the bias of hot deck imputation. Also, as a way of accounting for imputation uncertainty through multiple imputation, we base our method on Lo's (1988) finite population Bayesian bootstrap.

Results from simulations show that our method performs at least as well as competing methods for the estimation of means and confidence intervals, especially given a larger sample size and nonlinear relationships among the variables.

Key words: Kernel imputation; Finite population Bayesian bootstrap; Hot deck imputation; Boundary bias

References

Lo, A. Y. (1988). A Bayesian bootstrap for a finite population. *The annals of statistics*, **16**, 1684-1695.

[1]Department of Statistics, Stockholm University, Sweden

Finite Population Stratification Algorithm

Aleksandras Plikusas[1]

A simple and adaptive empirical stratification algorithm is presented. The algorithm depends on one parameter, and by choosing the value of this parameter we may stratify the estimation domains in a way which is close to optimal (see Krapavickaitė and Plikusas (2010)).

Consider the survey population $\mathcal{U} = \{u_1, \ldots, u_N\}$ of N elements, and a study variable y, defined on this population, taking nonnegative values $y_1, \ldots, y_N$. We are interested in the estimation of the finite population total using a simple random stratified sampling. The problem is to find the stratification boundaries that minimize the variance of the Horvitz-Thompson estimator of the total $t = \sum_{k=1}^{N} y_k$. Assume that the values of the study variable are sorted in increasing order. We also assume that the number of strata H and sample size n are given in advance. The proposed stratification algorithm is as follows. From equations

$$\sum_{k=1}^{k_1} y_k^{\alpha} = \sum_{k=k_1+1}^{k_2} y_k^{\alpha} = \sum_{k=k_{H-1}+1}^{N} y_k^{\alpha} \tag{1}$$

find indices $k_1, k_2, \ldots, k_{H-1}$, and set stratum boundaries $b_0 = 0$, $b_1 = y_{k_1}$, $b_2 = y_{k_2}, \ldots, b_{H-1} = y_{k_{H-1}}$, $b_H = y_N$. The stratification boundaries found depend on the chosen α which may be determined by simulation.

The proposed stratification method is compared by simulation with cumulative square root method of Dalenius and Hodges (1959), geometric stratification (see Gunning and Horgan (2004)), and Lavallee-Hidiroglou (1988) method.

References

Dalenius T, Hodges, J.L. (1959). Minimum variance stratification, Journal of the American Statistical Association, 285, 88-101.

Gunning, P., Horgan, J. M. (2004). A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations, Survey Methodology, 2, 159-166.

Krapavickaitė, D., Plikusas, A. (2010). Some choices of a specific sampling design. In: Official Statistics, Methodology and applications in honour of Daniel Thorburn, Stockholm University, 79-92.

P. Lavallee, P., Hidiroglou, M. A. (1988). On the stratification of skewed populations, Survey Methodology, 33-43.

[1]Vilnius University, Lithuania

UKRAINIAN BUSINESS TENDENCY SURVEYS: PROBLEMS & PERSPECTIVES

Maryna Pugachova[1]

Business Tendency Surveys (BTS) conducted in Ukraine with the quarterly periodicity since 1997. Some problems arisen at the beginning of works (absence of business register, absence of comprehension and perception of such survey as important source of additional statistical information and others).

BTS cover 6 branches of economy (industry, construction, retail trade, transport, agriculture and non-financial services sector) now.

The issues concerning samples by sectors, using of imputation techniques, different blocks of questions in BTS questionnaires, some tendencies in each sector and presentation of the results for consumers will be viewed.

The method of constructing of the composite (synthetic) indicators for different sectors of economy on the basis of Ukrainian BTS data will be presented (whose experience we used and how we choose the BTS indicators for the synthetic indicators, how we construct the leading and coincident synthetic indicators).

Last part of report contains the Ukrainian results of application of Ifo-Institute segmentations method for detecting and forecasting of cycles turning points as one way of BTS data utilization.

[1] Scientific & Technical Complex for Statistical Research
of the State Statistics Committee of Ukraine

ESTIMATION OF QUADRATIC FINITE POPULATION FUNCTIONS

Dalius Pumputis[1]

Consider a finite population $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$ of N elements. Let $y^{(k)} : y_1^{(k)}, y_2^{(k)}, \ldots, y_N^{(k)}$, $k = 1, 2, \ldots J$, be J study variables defined on the population $\mathcal{U}$. The values of the variables $y^{(k)}$, $k = 1, 2, \ldots J$, are known only for sampled population elements. We assume also that there are available J known auxiliary variables $x^{(k)}$, $k = 1, 2, \ldots J$, with values $x_1^{(k)}, x_2^{(k)}, \ldots, x_N^{(k)}$. Denote $\mathbf{y}_i = (y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(J)})$ and $\mathbf{x}_i = (x_i^{(1)}, y_i^{(2)}, \ldots, x_i^{(J)})$.

We are interested in the estimation of the quadratic finite population function

$$T = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \phi(\mathbf{y}_i, \mathbf{y}_j)$$

under the general sampling design. Here $\phi(\cdot, \cdot)$ is a symmetric function (a kernel of degree 2 for a U statistic). As the special cases of the function T, are well known parameters, as the finite population variance, the covariance, the variance of the Horvitz-Thompson estimator. The parameter T can be viewed as a total of a certain variable. So, some calibration methods can be employed to derive the estimators of T.

In the case of none auxiliary information, one can take the standard Horvitz-Thompson type estimator $\widehat{T}_{HT}$. In the paper of Sitter and Wu (2002) there are introduced model-calibrated estimators of the quadratic finite population function T.

Using Deville and Särndal's (1992) calibration technique, we derive first a corresponding calibrated estimator $\widehat{T}_{DS}$. Then, using Farrell and Singh's (2002, 2003) penalized distance measure, we introduce a penalized estimator $\widehat{T}_P$ of quadratic function T. It is shown that the mean square error of estimator $\widehat{T}_P$ is approximately equal to variance of $\widehat{T}_{DS}$ multiplied by $1\backslash(1+\varphi^2)$, here φ is the penalty parameter. If $\varphi \to \infty$, then mean square error of estimator $\widehat{T}_P$ is decreasing, however the bias $B(\widehat{T}_P)$ is extremely increasing. So, we are looking for the possibilities how to improve this estimator. Solving this problem we developed a new distance measure and a new calibration equation. Minimization of this distance measure subject to a new calibration equation leads to the estimator

$$\widehat{T}^* = \alpha\widehat{T}_{HT} + \beta\widehat{T}_{DS} + \gamma\widehat{T}_P, \quad \alpha + \beta + \gamma = 1,$$

which is the linear combination of estimators discussed above. Limited simulation results show that for some sets of α, β and γ the new estimator $\widehat{T}^*$ is more accurate than penalized estimator $\widehat{T}_P$. Of course, the optimal values of α, β and γ may be derived and used in $\widehat{T}^*$.

References

Deville, J. C.; Särndal, C. E.: Calibration estimators in survey sampling. Journal of the American Statistical Association. 87, 376–382 (1992)

Farrell, P.; Singh, S.: Penalized chi square distance function in survey sampling. ASA Proceedings. 963–968 (2002)

Singh, S.: On Farrell and Singh's penalized chi square distance functions in survey sampling. SCC Proceedings. 173–178 (2003)

Sitter, R. R.; Wu, C.: Efficient estimation of quadratic finite population functions in the presence of auxiliary information. Journal of the American Statistical Association. 97(458), 535–543 (2002)

[1]Vilnius Pedagogical University, Lithuania

SOME ASPECTS OF SAMPLING ERROR ESTIMATION IN OFFICIAL STATISTICS

Rudi Seljak[1], Petra Blažič[2]
Statistical Office of the Republic of Slovenia

ABSTRACT

Key words: sampling error, quality assessment, metadata driven application

Although in the modern quality assessment framework precision of the statistical results, obtained from the data of the sampling survey, is considered just as one of the several quality dimensions, it still remains one of the key aspects when the reliability of these results is judged. It is therefore of great importance that precision (assessed mostly through the sampling error estimation) is regularly estimated and adequately provided to the users. Unfortunately in the case of modern official statistics production when short timeliness and wide exhaustiveness of the disseminated results is more and more demanded, regular assessment of precision is often neglected and consequently the point estimates are disseminated without any information on the precision Results with the lower precision can consequently be used by users in order to draw conclusions which could be misleading.

At the Statistical Office of the Republic of Slovenia revision of the system of the sampling error estimation as well as the system of the presentation of these errors to the users started a few years ago. The main goal of the revision was to set up a system which would enable quick and effective sampling error estimation and to define rules; how the information about precision should be provided to the users in a clear and transparent way. In the paper we describe some of the dilemmas and trade-offs regarding the sampling error estimation, which we faced when the revision was carried out. We also shortly describe the general application for sampling error estimation, developed in the framework of the revision. The application is build on the bases of so called metadata driven principle, meaning that there is one general program code which is then for the particular survey parameterized through the (process) metadata tables.

[1] Rudi Seljak, Statistical Office of the Republic of Slovenia, Vožarski pot 12, 1000 Ljubljana, Slovenia, rudi.seljak@gov.si, phone: +386 1 2415 294
[2] Petra Blažič, Statistical Office of the Republic of Slovenia, Vožarski pot 12, 1000 Ljubljana, Slovenia, petra.blazic@gov.si, phone: +386 1 2415 126

Finite mixtures analysis in survey sampling problems

Artem Shcherbina
National Taras Shevchenko University of Kyiv, Ukraine
E-mail: artshcherbina@gmail.com

Observations from mixtures of different subpopulations are frequent in biological and sociological studies. In this communication we consider the case, when the observations are taken from a set of groups which contain subjects belonging to different subpopulations. Proportion of each population in a group is known and can vary from group to group. Our aim is to estimate the means of an observed variable for subjects belonging to each subpopulation.

Such problems arise in analysis of sociological surveys data concerning so called "sensitive questions", e.g. questions on drugs usage or cheating on school tests and so on. Anonymous surveying is usually used to avoid inadequate answers on such questions. On the other hand, it is interesting to compare the obtained anonymous information on the proportion, say of cheaters in different groups of anonymous respondents to open information on these individual features, such as age, school marks, gender, etc. In this example one considers two subpopulations of cheaters and non-cheaters and estimate mean characteristics over these subpopulations.

Another example is an analysis of genetic and phenotype information in genomic imprinting studies.

In the communication we consider some nonparametric estimates to the subpopulation means such as weighted means with minimax and adaptive weights, polynomial least squares estimates. Finite sample properties and asymptotic behavior of these estimates are discussed. They are compared to maximum likelihood estimates for some parametric submodels.

ESTIMATION FOR DOMAINS AND SMALL AREAS FOR BUSINESS SURVEYS

Milda Šličkutė-Šeštokienė [1]

Statistics Lithuania has the full range of labour statistics that meet the timeliness and demands of Eurostat and national needs. The challenge is to keep this quality and timeliness and to publish even more detailed information and at the same time spare costs.

Users need more and more statistical information and at the same time respondents want to get less and less questionnaires. That enforces Statistics Lithuania to seek for new methods for estimation of statistical information required. Administrative sources (e.g. data of Social Insurance) plays a significant role trying to increase the quality of the results, to diminish the burden for the respondents and to spare the costs.

This contribution is devoted to the usage of administrative sources at the stage of estimation for domains, including small areas. Design-based estimators (Horvitz-Thompson and GREG) versus model-based (Synthetic and EBLUP) estimators are analyzed for estimation of totals and ratios of two totals. Simulation study is accomplished using real data of Lithuanian Quarterly Survey on Earnings. The effect of model choice is examined and planned domains versus unplanned domains are analyzed.

The main issue of Lithuanian Quarterly Survey on Earnings is multi-purpose of the survey so that a lot of variables are included for a different purposes and also a lot of breakdowns are required. This multi-purpose of the survey cause a lot of inconvenience at the estimation stage, because not only best estimators should be found for different variables but also coherence between different variables should be preserved.

Keywords: Survey sampling, generalized regression estimator, GREG, ratio estimator, administrative sources, auxiliary information.

References

Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, **29**, 33-44.

Lehtonen, R., Särndal, C.-E. and Veijanen, A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, **7**, 649-674.

Lehtonen, R., Pahkinen E.J. (2004). *Practical Methods for design and analysis of Complex Surveys*. Second Edition. Chichester: John Willey & Sons.

Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model-assisted Survey Sampling*. New York: Springer-Verlag.

[1] Statistics Lithuania, Lithuania

3 RUSSIAN SURVEYS: COMPARATIVE ANALYSIS

Vladimir V. Ulyanov[1]

3 sociological surveys were carried out in Russia in 2007, 2009 and 2011 with appr. 7000, 2000 and 1500 randomly-selected adults (18+) respectively from all federal districts. The questionnaires of the surveys with around 50 questions (appr. 170 subquestions) were almost identical.

Applying factor analysis, multidimensional scaling and cluster analysis we construct a semantic space in order to model a process of Russian society consolidation. The profiles of corresponding typological groups are described. We suggest a composite index of social consolidation. In the talk we discuss the methodological aspects of the surveys analysis itself.

We consider as well an attempt to apply qualitative comparative analysis suggested by Charles Ragin (see e.g. Benoit Rihoux (2006)).

References

[1] *Benoit Rihoux* Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research. International Sociology. 21, 679-706 (2006)

[1]Moscow State University, Russia

ADJUSTMENT FOR MEASUREMENT ERRORS: SIMULATION STUDY

Maria Valaste[1] and Risto Lehtonen[1,2]

In this paper we investigate three methods for adjustment for measurement errors in surveys. The methods are Maximum Likelihood, Multiple Imputation and Regression Calibration. These methods requires information obtained from validation study. If the model is correct the Maximum Likelihood (ML) estimators of the parameters will have nice properties. The ML estimators of the parameters will be consistent, asymptotically normal and have the smallest asymptotic MSE among all regular estimators. (Messer and Natarajan, 2008). Because of these properties of ML estimators and motivation by some earlier publications (Spiegelman et al., 2000; Messer and Natarajan, 2008, e.g.) the ML estimator of parameter will set the gold standard for comparison.

An interesting approach dealing measurement errors is multiple-imputation for measurement errors (Cole et al., 2006; Padilla et al., 2009). In MIME approach measurement errors are treated as a missing data problem. Regression Calibration method is widely applied and studied (Rosner et al., 1989; Kuha, 1994; Spiegelman et al., 2000, 2001; Messer and Natarajan, 2008). Regression calibration is a statistical method for adjusting point and interval estimates for bias due to measurement error.

An extensive Monte Carlo simulation studies is conducted to test and have experience of the three methods: MI, ML and RC. The properties of the MI and ML approaches was investigated and preliminary results was introduced in papers by Valaste et al. (2010a,b). Artificial data will be generated based on Finnish ECHP data of years 1996 and 2000, where the variables of interest such as income are measured both by interview and by administrative registers.

References

Cole, R. S., H. Chu, and S. Greenland (2006). Multiple-imputation for measurement-error correction. *International Journal of Epidemiology 35*, 1074-1081.

Kuha, J. (1994). Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Statistics in Medicine 13*, 1135-1148.

Messer, K. and L. Natarajan (2008). Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. *Statistics in Medicine 27*, 6332-6350.

Padilla, M. A., J. Divers, L. K. Vaughan, and D. B. Allison (2009). Multiple imputation to correct for measurement error in admixture estimates in genetic structured association testing. *Human Heredity 68*, 65-72.

Rosner, B., W. C. Willet, and D. Spiegelman (1989). Corrections of logistic regression relative risk estimates for systematic within-person measurement error. *Statistics in Medicine 8*, 1051-1069.

Spiegelman, D., R. J. Carroll, and V. Kipnis (2001). Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Statistics in Medicine 20*, 139-160.

Spiegelman, D., B. Rosner, and R. Logan (2000). Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association 95* (449), 51-61.

Valaste, M., R. Lehtonen, and K. Vehkalahti (2010a). Measurement errors in surveys: A multiple imputation approach. In International Conference on Indicators and Survey Methodology, 24.-26.2.2010, Vienna, pp. 57-58.

Valaste, M., R. Lehtonen, and K. Vehkalahti (2010b). Multiple imputation for measurement error correction in survey data. In Q2010 European Conference on Quality in Official Statistics, May 4-6 2010, Helsinki, pp. 89.

1 Social Insurance Institution, Finland
[2] University of Helsinki, Finland

RECURRENT OPTIMAL ESTIMATORS UNDER ROTATION CASCADE SCHEME THROUGH CHEBYSHEV POLYNOMIALS

Jacek Wesołowski *

Patterson, in his seminal paper (Patterson, 1950), derived recursion for optimal estimator of the mean on a given occasion for rotation schemes with no "holes". In such schemes a unit, leaving the sample on a certain occasion, never returns. The recursion in case of rotation patterns with "holes" had not been available for many years, in spite of the fact that such schemes were used in regular surveys.

Under cascade rotation patterns with singleton "holes" the recursion was derived recently in Kowalski (2009). Even more recently a stationary version of the problem for cascade rotation patterns with arbitrary "holes" has been settled in Kowalski and Wesołowski (2011). Unfortunately, the solution is based on two unpleasant assumptions. The first deals with localization of roots of a rather complicated polynomial. The second concerns unique solution of a complex system of linear equations. However, numerical experiments suggested that both these assumptions may be universally satisfied.

Using a new approach involving Chebyshev polynomials we will show that the first assumption is universally satisfied. The problem if the second assumption is also universally satisfied remains open. But in some cases it can be proved analytically. In particular, following Wesołowski (2010), we will present a complete solution of the recurrence problem in the case of 110011 rotation pattern, which is typical for the Labour Force Survey.

References

1. Kowalski, J., Optimal estimation in rotation patterns. *J. Statist. Plan. Infer.* 139(4) (2009), 2429-2436.

2. Kowalski, J., Wesołowski, J., Recurrence optimal estimators for rotation cascade patterns with holes (2011) - in preparation.

3. Patterson, H.D., Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc., B* 12 (1950), 241-255.

4. Wesołowski, J., Recursive optimal estimation in Szarkowski rotation scheme. *Statist. Transit.* 11(2) (2010), 267-285.

*Central Statistical Office, Warsaw, Poland

Program

	Monday	Tuesday	Wednesday	Thursday	Friday
9.00–10.30	**Welcome** *Chair* Traat I **G. Kulldorff** **D. Thorburn**	*Chair* Krapavickaite D **Jean-Claude Deville 3**	*Chair* Malmdin J **Giovanna Ranalli 3**	*Chair* Thorburn D **Steven Thompson 1**	*Chair* Lapins J **Imbi Traat**
	Jean-Claude Deville 1	**Anders Norberg**	**Natalie Shlomo**	**Steven Thompson 2**	**Steven Thompson 3**
	Refreshments	Refreshments	Refreshments	Refreshments	Refreshments
11.00–12.30	*Chair* Liberts M **Jean-Claude Deville 2**	*Chair* Lehtonen R **Giovanna Ranalli 1**	*Chair* Thompson S **Carl-Erik Särndal**	*Chair* Deville JC **Lennart Bondesson**	*Chair* Ranalli G **Risto Lehtonen**
	Ineke Stoop	**Giovanna Ranalli 2**	*Contr Session 3* **Small Areas** Slickuté-Sestokiene, M Nekrasaite-Liege, V Magnusson, M	*Contr Session 5* **Household surveys** Aru, J Bokun, N Krapavickaite, D	*Closing session* **Gunnar Kulldorff**
	Lunch	Lunch	Lunch	Lunch	Lunch
13.30–15.15	*Contr Session 1* *Chair* Shlomo N **Methodology in Practice** Orlova, J. Liberts, M Fisenko, A	**Excursion**	*Contr Session 4* *Chair* Bokun N **Sampling 1** Wesolowski, J Shcherbina, A Olofsson, J	*Contr Session 6* *Chair* Plikusas A **Quality and Nonsampling Errors** Seljak, R, Blazic, P Iashchenko, L Honchar, O	
	Pugachova, M Bobrova, A Larchenko, A		Koyuncu, N Gladun, O Ulyanov, V	Valaste, M Laaksonen, S Bueno, E	
	Refreshments		Refreshments	Refreshments	
15.45–17.00	*Contr Session 2* *Chair* Särndal C.-E **Calibration** Plikusas, A Pumputis, D Laitila, T Ilves, M		**Poster session** *Chair* Laitila T Pettersson, N Buceniece, B Bueno, E Budkina, N Bondarenko, I	*Contr Session 7* *Chair* Wesolowski J **Sampling 2** Berger, Y Grafström, A Ciginas, A Lumiste, K	
	Dinner		Dinner	Conf Dinner	

Participants

Family name	First name	Organisation/Institution/Company	E-mail address
Arnoldsson	Göran	Department of Statistics, Umeå University	Goran.Arnoldsson@stat.umu.se
Aru	Julia	Statistics Estonia	julia.aru@stat.ee
Berger	Yves	University of Southampton	Y.G.Berger@soton.ac.uk
Blazic	Petra	Statistical Office of the Republic of Slovenia	petra.blazic@gov.si
Bobrova	Anastacia	Institute of economy of National Academy of Science	nastassiabobrova@mail.ru
Bokun	Natalia	Belorussian state economic university	nataliabokun@rambler.ru
Bondarenko	Yana	Oles` Honchar National University	yanabondarenko@ua.fm
Bondesson	Lennart	Dept of Mathematics and Mathematical Statistics	lennart.bondesson@math.umu.se
Buceniece	Baiba	Central Statistical Bureau of Latvia	baiba.buceniece@csb.gov.lv
Budkina	Natalja	University of Latvia	budkinanat@gmail.com
Bueno Castellanos	Edgar Mauricio	Colombian Observatory of Science and Technology -OCyT-	embuenoc@ocyt.org.co
Ciginas	Andrius	Vilnius University	andrius.ciginas@mif.vu.lt
Deville	jean-claude	Ecole Nationale de la Statistique et de l'Analyse de l'Information/Crest	deville@ensae.fr
Fisenko	Andris	Central Statistical Bureau of Latvia	andris.fisenko@csb.gov.lv
Gladun	Oleksandr	M.V.Ptukha Institute for Demography and Social Researches	gladun@i.com.ua
Grafström	Anton	Swedish University of Agricultural Sciences	anton.grafstrom@slu.se
Gulbina	Ilze	TNS Latvia	liene.findleja@tns.lv
Hellberg	Olivia	Stockholm University	olivia.hellberg@stat.su.se
Honchar	Oksana	National Academy Statistics, Accounting and Audit	ohonchar@list.ru
Iashchenko	Liudmyla	Scientific and Technical Complex of Statistical Research	lud_ya@mail.ru
Ilves	Maiki	Örebro University	maiki.ilves@oru.se
Koyuncu	Nursel	Hacettepe University	nkoyuncu@hacettepe.edu.tr
Krapavickaite	Danute	Vilnius Gediminas Techn. University, Vilnius Univ. Inst. of Mathem. and Inform.	Danute.Krapavickaite@mii.vu.lt
Kulldorff	Gunnar	University of Umeå	gunnar@matstat.umu.se
Laaksonen	Seppo	University of Helsinki	Seppo.Laaksonen@Helsinki.Fi
Laitila	Thomas	Örebro university and Statistics Sweden	thomas.laitila@oru.se
Lapins	Janis	Bank of Latvia	Janis.Lapins@bank.lv
Larchenko	Anna	Belarus State Economic University	hanna-larchenko@mail.ru
Lehtonen	Risto	University of Helsinki	risto.lehtonen@helsinki.fi
Liberts	Martins	Central Statistical Bureau of Latvia	martins.liberts@gmail.com
Lumiste	Kaur	Estonian-Swedish Mental Health and Suicidology Institute	kaur.lumiste@eesti.ee

Magnusson	Måns	Stockholm University and Swedish Institute for Communicable Disease Control	mons.magnusson@gmail.com
Malmdin	Joakim	Statistics Sweden	joakim.malmdin@scb.se
Nekrasaite-Liege	Vilma	Vilnius Gediminas technical University	nekrasaite.vilma@gmail.com
Norberg	Anders	Statistics Sweden	anders.norberg@scb.se
Olofsson	Jens	Örebro University	jens.olofsson@gmail.com
Orlova	Yulia	Belarus State Economic University	orlova-julia-gen@mail.ru
Oyesetan	Prince Waheed	International Muslim College	oye_pr_2001@hotmail.com
Pettersson	Nicklas	Stockholm University, Department of Statistics	nicklas.pettersson@stat.su.se
Plikusas	Aleksandras	Vilnius University	Aleksandras.Plikusas@mii.vu.lt
Priedola	Inta	TNS Latvia	inta.priedola@tns.lv
Pugachova	Maryna	Scientific & Technical Complex for Statistical Research	maryni@ukr.net
Pumputis	Dalius	Vilnius Pedagogical University	dalius.pumputis@vpu.lt
Ranalli	Maria Giovanna	Universita' degli Studi di Perugia	giovanna.ranalli@stat.unipg.it
Rozora	Natalija	Nielsen	rozora@ukr.net
Seljak	Rudi	Statistical Office of the Republic of Slovenia	rudi.seljak@gov.si
Shcherbina	Artem	Taras Shevchenko National University of Kyiv	artshcherbina@gmail.com
Shlomo	Natalie	Southampton Statistical Sciences Research Institute	N.Shlomo@soton.ac.uk
Slickute-Sestokiene	Milda	Statistics Lithuania	milda.slickute@stat.gov.lt
Sõstra	Kaja	Statistics Estonia	kaja.sostra@stat.ee
Stoltze	Peter	Statistics Denmark	psl@dst.dk
Stoop	Ineke	The Netherlands Institute for Social Research/SCP	i.stoop@scp.nl
Särndal	Carl-Erik	Statistics Sweden and Orebro University	carl.sarndal@telia.com
Teneng	Dean	Statistics Institute, Tartu University	dean_teneng@yahoo.com
Thompson	Steven	Simon Fraser University	thompson@fas.sfu.ca
Thorburn	Daniel	Stockholm University	Daniel.thorburn@stat.su.se
Traat	Imbi	University of Tartu	imbi.traat@ut.ee
Ulyanov	Vladimir	Faculty of Computational Mathematics and Cybernetics, Moscow State University	vulyan@gmail.com
Valaste	Maria	The Social Insurance Institution of Finland	maria.valaste@helsinki.fi
Wesolowski	Jacek	Glowny Urzad Statystyczny (Central Statistical Office)	wesolo@mini.pw.edu.pl