

ESTIMATION FOR DOMAINS AND SMALL AREAS

O. Vasylyk¹

¹ National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
e-mail: vasylyk@matan.kpi.ua

Abstract

Estimation for domains is concerned with the development of statistical procedures for producing efficient estimates for population parameters, such as totals, means, proportions, medians, quantiles, percentiles for population subgroups called domains. Domains are defined by the cross-classification of geographical districts by social, economic, demographic characteristics. Small area estimation (SAE) means estimation for domains whose sample size is small or very small (even zero).

There are several classifications of SAE methods, in particular, direct and indirect methods, design-based methods and model-based methods, area-level models and unit-level models. Direct methods use only domain-specific data. Indirect methods borrow information from all the data. Design-based methods often use a model for the construction of the estimators (model assisted), but the bias, variance and other properties of the estimators are evaluated under the randomization (design-based) distribution. Model-based methods usually condition on the selected sample, and the inference is with respect to the underlying model. Area-level models relate small area direct estimators to area specific covariates. Such models are necessary if unit (or element) level data are not available. Unit-level models relate the unit values of a study variable to unit-specific covariates.

In the lecture, we shall consider classical and new approaches to estimation for domains and small areas. Also a brief overview of SAE software will be presented and some examples of computing the estimates will be given.

Keywords: design-based methods, domains, model-based methods, small area estimation.

References

- Ghosh, M., Rao, J.N.K. (1994). Small area estimation: an appraisal. *Statistical Science*, **9**, 55-93.
- Lehtonen, R., Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester: John Wiley & Sons.
- Lehtonen, R., Veijanen, A. (2009). *Design-based methods of estimation for domains and small areas*. In *Sample Surveys: Inference and Analysis* (D. Pfeffermann and C. R. Rao, eds.). *Handbook of Statistics*, **29B**, 219-249. North-Holland, Amsterdam.
- Lehtonen, R., Veijanen, A. (2020). Hybrid calibration methods for small domain estimation. *Statistics and Applications*. **XVII**. 201-235.
- Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40-68.
- Pratesi, M., Giusti, C. (2015) Small area estimation I. *1st EMOS Spring School*, Trier, Pisa, Manchester, Luxembourg, 23-27 March 2015.
- Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

ОЦЕНИВАНИЕ ДЛЯ ДОМЕНОВ И МАЛЫХ ОБЛАСТЕЙ

О. Васылык

Национальный технический университет Украины “Киевский политехнический институт имени Игоря Сикорского”, Украина

e-mail: vasylyk@matan.kpi.ua

Аннотация

Оценивание для доменов связано с разработкой статистических процедур для получения эффективных оценок параметров, таких как суммарные значения, средние значения, пропорции, медианы, квантили, процентиля для подгрупп популяции (генеральной совокупности), называемых доменами. Домены определяются перекрестной классификацией географических районов по социальным, экономическим и демографическим характеристикам. Оценивание для малых областей (SAE) означает оценивание для доменов, размер выборки из которых мал или очень мал (даже может быть равен нулю).

Существует несколько классификаций методов оценивания для малых областей, в частности, прямые и косвенные методы оценивания; методы, основанные на дизайне выборки и методы, основанные на моделях; модели уровня области и модели уровня единицы. Прямые методы используют только данные, относящиеся к домену. Косвенные методы заимствуют информацию из всех данных. Методы, основанные на дизайне, часто используют модель для построения оценок, но смещение, дисперсия и другие свойства оценок оцениваются в рамках рандомизационного (основанного на дизайне) распределения. Методы, основанные на моделях, обычно зависят от полученной выборки, а выводы делаются относительно базовой модели. Модели уровня области связывают прямые оценки для малых областей с ковариатами, характерными для этих областей. Такие модели необходимы, если данные на уровне единицы (элемента) недоступны. Модели уровня единицы связывают значения исследуемой переменной для элемента с ковариатами, специфичными для элемента.

В лекции мы рассмотрим классические и новые подходы к оцениванию для доменов и малых областей. Также будет представлен краткий обзор программного обеспечения для SAE и приведены некоторые примеры расчета оценок.

Ключевые слова: домены; методы, основанные на дизайне выборки; методы, основанные на моделях; оценивание для малых областей.