HUNGARIAN CENTRAL STATISTICAL OFFICE

About the practice of selecting data suppliers for business statistics surveys conducted by the HCSO

Contents

Legal background to obligation to provide data 3
Observation units for business statistics surveys 3
Information used for selecting data suppliers for business statistics surveys 4
Full-scope data collections and samplings 4
Methodology of sampling used in business statistics surveys 5
Modifying obligation to provide data 6

Contact details

Legal background to obligation to provide data

The supply of statistical data for business statistics surveys is based on the **Government Decree** on the **National Statistical Survey Programme, issued upon the authorisation of Act on Statistics** (Act CLV of 2016 on Official Statistics), having regard to Regulation (EU) No. 549/2013 and Commission Decision 98/715/EC.

The supply of statistical data to the Hungarian Central Statistical Office (HCSO) is compulsory based on Section 26 (1) of Act on Statistics, the data serve solely statistical purposes[1]. Refusing data supply, providing false information and being behind in responding may imply legal action.

Observation units for business statistics surveys

The aim of business statistics surveys is to produce high-quality statistical information on the state of and changes in the economy, the society and the environment by observing establishments[2] performing activities relevant to the data collections. To reach this goal, in the frame of business statistics surveys we collect data on establishments, basically they are the observation units for these surveys. In the practice of business statistics surveys by the HCSO, we observe in addition to establishments other statistical units, too, according to the characteristics of the different statistical surveys (e.g. shops in retail trade statistics, private holdings in agricultural statistics, non-profit organisations in labour statistics, products and services in price statistics). Selecting data suppliers for surveys of such special observation units may be different from what is discussed in this document.[3]

To define the observation units for surveys in a given period **we use a *sample frame*,** which contains the list of potential data suppliers and their characteristics, based on the thoroughly gathered best-ever information. The HCSO produces the best-ever information on the basis of statistical registers (which are at the same time the sources of sample frames) maintained using many statistical surveys and administrative data sources. In the field of business statistics surveys one of the most important statistical registers is the Business Register (BR). For business statistics surveys the HCSO uses besides the BR other statistical registers (satellite registers), more information on which is available from meta-information describing the data sources of the Office.[4]

Every statistical register is to contain the most up-to-date information possible on the different statistical units. However, **there may be differences between the information stored in statistical registers and the information current at the time of producing sample frames.** The reason behind is that because of the characteristics of administrative processes, registers cannot follow up every moment the changes in the phenomenon or population under review, this new information describing the individual or the population will be included in them after some time only. Therefore, the HCSO typically uses many data sources to maintain statistical registers and be able to define the scope of data suppliers of business statistics surveys based on the most up-to-date information possible.

1 More details on the use of data for statistical purposes can be found under the information for data providers on confidentiality: http://www.ksh.hu/information_on_confidentialty_for_data_providers

2 Here and hereinafter we use establishments as the synonym for observation units for surveys in general.

3 Detailed information on observation units, data production methodology, data sources used, etc. for each statistical domain is available in annual descriptions of the statistical domains: http://www.ksh.hu/apps/meta.menu?p_lang=EN&p_menu_id=110&p_session_id=32748467

4 Information on data sources and statistical registers of the HCSO: http://www.ksh.hu/apps/meta.menu?p_lang=EN&p_menu_id=1410&p_session_id=37889847

Information used for selecting data suppliers for business statistics surveys

The HCSO prepares a statistical principal activity code for each establishment every year. **The principal activity of a statistical unit is the activity which contributes most to the total value added of the particular statistical unit.** The statistical principal activity code of an establishment is automatically equal to its administrative principal activity code by default. However, the HCSO defines the code of the establishments on which the HCSO has more up-to-date information from other data sources by a 'statistical principal activity settings algorithm'. Namely, the HCSO uses actual activity instead of administrative principal activity as an attribute for the selection.

Besides the statistical principal activity, information on the size of establishments is also highly important in selecting data suppliers for business statistics surveys. **To have this information we record in the BR the number of active employees and revenue, based on which we group the establishments into categories of number of employees and revenue.** Setting the categories of number of employees and revenue is based on data reported to the HCSO in a given year as well as on administrative data in the given and the previous year, according to an algorithm built on a pre-defined priority process.

The data used for the selection of data suppliers are unchanged during the data collection year. So a data supplier may change category of number of employees or revenue due to changes in the number of its employees or its revenue over the year, or its administrative principal activity may change, however, this does not alter its statistical principal activity or category of number of employees valid at the time of the selection. Naturally, the HCSO stores the reported changes, and works with the most up-to-date information available at the time of the selection for the next year.

The HCSO takes into account other characteristics as well at the time of defining the basis of the selection in some areas of business statistics because of the special need of the statistical domain: e.g. land area and livestock are observed in agricultural statistics, the number and capacity of means of transport (goods motor vehicles, tractors, etc.) in road goods transport statistics, exports and imports in external trade statistics and secondary activities in services producer price statistics.[5]

Full-scope data collections and samplings

A significant part of business statistics surveys are full scope (census), a smaller part are based on sampling. The aim of a full-scope statistical data collection is to observe all establishments in the target population, i.e. we observe all units in the survey frame covering the target population as much as possible (the target population of a data collection can be e.g. every establishment in particular industries and in particular categories of number of employees out of the establishments active in a particular area in the reference period). In the case of a sample survey we select a part of the survey frame of the target population so that it should represent the target population (probability sampling).

[5] Detailed information on observation units, data production methodology, data sources used, etc. for each statistical domain is available in annual descrip - tions of the statistical domains: http://www.ksh.hu/apps/meta.menu?p_lang=EN&p_menu_id=110&p_session_id=32748467

Methodology of sampling used in business statistics surveys

The main aspect of producing samples for statistical purposes is to be able to make high-quality estimates on the characteristics of the target population of a statistical survey (e.g. industrial production value, average number of people employed at an establishment, investment outlay, etc.) based on data collected from sample units. We make estimates from samples with the tools of mathematical statistics, according to methodologies recommended among others in official statistics of the European Statistical System and the relevant quality guidelines of the HCSO[6]. We regularly measure the quality of the produced estimates and describe it with different statistical indicators[7].

A prerequisite for the use of estimation tools and quality indicators is that the sample should be a probability sample[8]. The selection of the applied sampling procedure is influenced by many factors, among others the desired level of accuracy of the information to be produced, the accessibility and quality of adequate data sources, the availability of appropriate variables for making strata, estimation methods to be used and available budget resources.

The population of business statistics surveys is extremely heterogeneous in terms of the characteristics to be observed, therefore, **the sample is usually selected by *stratified* sampling,** i.e. establishments are divided into homogeneous groups according to their characteristics. The characteristics of the units differ to a lower extent from one another in these groups, so it is enough to select a sample with a smaller number of elements in order to reach the same level of accuracy of estimation. The most frequent variables for making strata are statistical principal activity code (based on TEÁOR'08[9]), categories of number of employees or revenue, and territorial characteristics (e.g. regions, counties or Budapest/the country).

Generally, establishments with large revenue and/or with a large number of active employees – due to the importance of the data and information they provide – are included in the sample of statistical surveys, since they contribute significantly to the quality of estimates to be made on the basis of this information.

One of the most crucial points of stratified sampling is to define the size of the sample for each stratum. The size of the sample for each stratum (out of the number of potential data suppliers, available according to the breakdown by the different strata, the number of data suppliers to be actually surveyed) is computed using optimal allocation, for which we mostly use tax data as auxiliary information, coming from administrative sources.

After defining the size of the sample for each stratum, the selection of the sample follows, i.e. the actual selection and specification of as many data suppliers as the defined numbers of elements.

When selecting samples, the HCSO uses sample co-ordination, the main purpose of which is to spread the response burden as fairly as possible. To this end, the Office continuously monitors the response burden placed on the establishments, and prefers as much as possible the establishments with less burden earlier on when selecting the sample.

The specific sampling elements are selected using a special variant of 'sequential simple random sampling'.[10] The method guarantees that within a particular stratum each establishment should be included in the sample with the same probability, at the same time makes it possible to take into consideration the response burden placed on the establishments and spread the total burden among them as evenly as possible.

Monitoring appropriately changes in the population, maintaining the selected sample and taking into consideration the response burden require that **we use *rotation* in the case of** certain **repeated surveys, that is, we periodically replace in the sample a part of the establishments selected to be surveyed.** For each sample survey, rotation is determined by the sampling plan of the given statistical survey. When determining the rules and possibilities of rotation, many aspects are to be considered, among others the degree of the response burden, the frequency of changes in and closures of establishments, the degree of attrition, etc. Whether these aspects can be considered depends significantly on the specific features of a particular survey. For instance, the size or frequency

6 Quality guidelines of the HCSO: http://www.ksh.hu/guidelines_policies_strategy
7 See the quality policy of the HCSO: http://www.ksh.hu/guidelines_policies_strategy
8 That is, each element of the population has a chance to be included in the sample, i.e. the probability of inclusion in the sample is known and positive for each element of the population.
9 See the standard industrial classification of all economic activities (TEÁOR'08): http://www.ksh.hu/teaor_eng_menu
10 We assign a random number, taken from the [0, 1) interval with uniform distribution, to each unit in the BR independently, taking into account the burden placed on the unit (the heavier the burden on the unit, the higher the random number assigned to it). Subsequently, in each stratum we put the units in the ascending order of the random numbers assigned to them, and select in each stratum the first n units to be data suppliers, in line with the sample size pre-defined for the particular stratum.

of rotation can be influenced by the stratification applied, so that in strata with a low number of elements there is less chance to replace the sampling elements.

There are business statistics surveys (e.g. price or goods statistics surveys) where the sample is selected by stratified concentrated sampling, i.e. the active establishments permanently present on the market and dominant in the sale of goods or services to be observed are selected, taking into account specific features of the different industries. Rotation cannot be used when selecting data suppliers for these surveys, whether establishments are included in or left out of the sample is basically influenced by economic processes.

Modifying obligation to provide data

Data suppliers are selected as a result of the procedure described above. **Modifying the group of establishments selected to be surveyed is not possible during the year.** The HCSO informs data suppliers on their obligation to provide data until 31 December each year or at the latest 30 days before the deadline for the compulsory data provision.

The HCSO needs a 'negative' response even if data suppliers – due to a change in their activities – deem that further on they do not fall within the group of data suppliers required by the HCSO to provide data. Therefore, sending back 'negative' questionnaires during the year remains compulsory.

kommunikacio@ksh.hu

Contact us!

Telephone: +36 (1) 345 6789