Current progress in development of indicators for data collection at Statistics Sweden

Gustaf Strandell
Statistics Sweden

The prevailing respondent climate has meant a high and dramatically increasing non response in the Swedish LFS. Attempts to slow the falling response rates through intensified interviewing work has in turn led to dramatically increased costs for Statistics Sweden and an unmanageable workload for Statistics Sweden's collection department. Prompted by the difficult situation, Statistics Sweden has initiated a set of projects to test and implement updates and alternatives to the currently used collection modes. Among other things, tests of web as an alternative to telephone, has been undertaken (in Statistics Sweden's Party Preference Survey) and the use of SMS in communication with the sample persons has been introduced on broad front. But the situation also places greater demands than ever on a data collection process that is characterized by a high degree of efficiency, where the allocation of resources in terms of interviewing hours and there usage over time, are well thought out and motivated.

In order to address the increasing difficulties in administering the data collection work, it is important that relevant measures is currently available, which makes it possible to monitor and control the data collection with efficiency, process and product quality in mind. In order to achieve efficiency and cost savings in the data collection, carefully planned strategies for the fieldwork are also needed, which ensures that relevant information from the collection work, including measures and indicators of the above kind, is recovered and used to control the process.

This mini paper describes a project running at Statistics Sweden during 2014-2015, with the task to develop and implement a set of indicators for ongoing monitoring and management of the data collection process for the Labor Force Survey (LFS). The project is described more fully in Swedish in SCB(2015). The indicators proposed by the project so far are based on relevant process data and provides a detailed view of the planned and realized input of resources as well as the result of the realized input of resources in data collection for the LFS. Indicators of this kind have to some extent been used for a long time at Statistic Sweden, but the need for an updated and expanded tool box as well as a more systematic use is great.

The project has also been given the task to develop a new and more cost-effective collection strategy for the LFS. Systematic usage of the new indicators should give improved visibility and transparency of the data

collection process, which leads to greater opportunities to set quantifiable targets for the process. Which is the minimum / maximum number of hours that should be spent by the interviewers during the current panel round on the group of sampled persons who were not contactable during the previous panel round? What response rate should be achieved in different groups after the first collection week? Along these lines the project has identified the need for a more uniform distribution of the workload for the interviewers during the period of time when the data collection is running for a particular LFS month.

Indicators

The conduction of a survey generates an amount of data which can be used to describe the field work. By processing the available process data, numerical information can be generated, which can be used as a basis for planning, evaluation and control of the data collection, both in the short term during data collection, and in more long term.

In recent years this issue has also attracted interest within the scientific community. An example of this is Kreuter (2013), which discusses how process data can be used in efforts to improve the sample surveys and highlights the problems that can arise in the analysis of process data. Chapter 9, authored by Jans, Sirkis and Morgan, contains a discussion of the use of process data for control of the data collection process in accordance with what is called Statistical Process Control (SPC). The work presented in this mini paper is not an attempt to push the research front on this field in any direction, but hopefully by implementing much needed updates of indicators and strategies for the Swedish LFS we will build a platform from which more advanced things can be done in the future.

The project has also chosen not to work (at this stage) with indicators that measure the quality of the estimates from the LFS. There are several reasons for this. The project group has identified the need to get the data collection process under control as most urgent. After that we can continue with the study of other more advanced indicators for the LFS, for example, indicators of representativity and the like. There are also other studies conducted at Statistics Sweden that are looking more closely into this.

The purpose of the indicators that the project proposes is to highlight and describe the data collection process in terms of planned work, work performed and the results of the work performed. The project has generated a somewhat large number of different indicators. Two examples are given below.

In the first example, we want to compare the realized resource input in the Labor Force Survey for April 2015 with the results achieved in terms of response rates. Since LFS is a panel survey, for all sample persons in the current panel round (except the newly recruited) there are information on the results of the data collection work of the previous panel round. In Table 1 below, the sample for the current panel round are divided into the groups new this round, interview last round, refusal last round, no contact last round, prevented to participate the last round and other reason for

nonresponse the last round. This grouping seems to be justified by the fact that the response rates are very different in the different groups, indicating also that the treatment of the sampled persons in the data collection probably should vary over the groups. The question about which goals should be set up for the different groups is still a work in progress. The goals presented in Table 1 are selected more or less ad hoc.

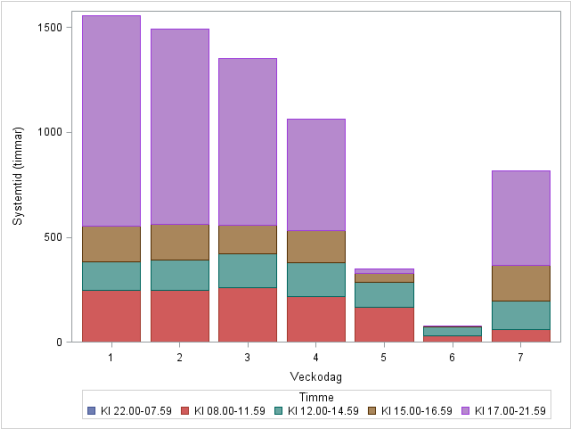
Table 1

			Results this round					System time		Contact time	
Result last round	**Number of sp**	**Response rate Goal**	**Response rate**	**Unable**	**Non contact**	**Refusal**	**Over cover.**	**Hours**	**Hours per sp**	**Hours**	**Hours per sp**
Total	25691		59.5	1.9	29.0	9.3	0.3	6753	0.3	4812	0.2
New	3663	**64**	57.1	1.5	29.6	11.0	0.9	1278	0.4	956	0.3
Interview	13518	**92**	86.1	0.7	11.1	2.1	0.1	3672	0.3	2269	0.2
Refusal	2165	**12**	11.3	0.9	31.0	56.7	0.1	324	0.2	286	0.1
Noncontact	5426	**27**	16.0	1.1	74.4	8.0	0.5	1317	0.2	1178	0.2
Unable	441	**5**	18.8	58.7	17.9	4.3	0.2	73	0.2	56	0.1
Other	478	**92**	74.5	1.3	17.8	6.5	.	89	0.2	67	0.1

All interviews for the LFS are carried out in SCB's interview system Windati. From Windati it is possible to read off the time span that the interviewers devote to each sample person. Table 1 provides this information aggregated in the variable System time. It is also possible from Windati to see for how much of the system time that the interviewing form is open. The system time during which the interview form is not open is called Contact time (and it is assumed that the interviewer use this time for contact trials). From a total resource perspective, we see that the groups interview last time and noncontacts last time are most expensive in terms of contact time (which of course follows from the fact that these groups are bigger than the rest). In the group non-contact last time the realized resource input gives a rather meager result.

The second example shows during which days of the week and during which times of the day that the system time has fallen out.

Figure 1



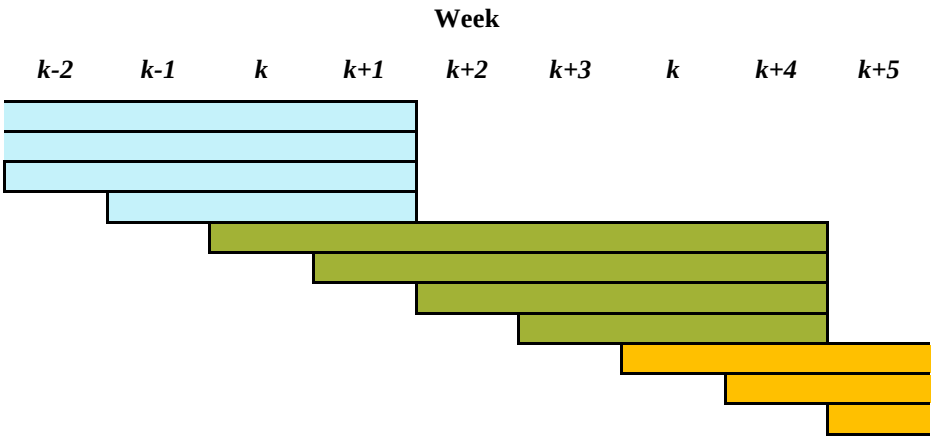
The interviewers work mostly in the evening between 17:00 and 21:59. Only a small amount of work is conducted on Fridays and Saturdays.

Strategies

For a number of established key indicators the project wants to quantify the goals against which the ongoing data collection should be monitored and evaluated. This is a work that has only just begun. The most important thing initially has been to follow the indicators systematically and study them in order to learn how the ordinary process is running and try to learn as much as possible for the future. Based on this study, realistic goals regarding for example use of resources and response rates can hopefully be quantified. It should also be clear from such a strategy what actions should be taken if the production targets are not reached. This requires further work.

However, along these lines the project has identified the need for a more uniform distribution of the workload for the interviewers during the period of time when the data collection is running for a particular LFS month. Each LFS-month consists of four or five reference weeks that start their field work Monday after the current reference week. For the data collection, this means therefore that a new field work starts every Monday throughout the year. Today the fieldwork for each reference week in the LFS-month is allowed to run until the end of the month. Figure 2 below illustrates how this strategy in practice means that we can have six parallel ongoing collection rounds at once (four for the current month and two for the following month).

Figure 2

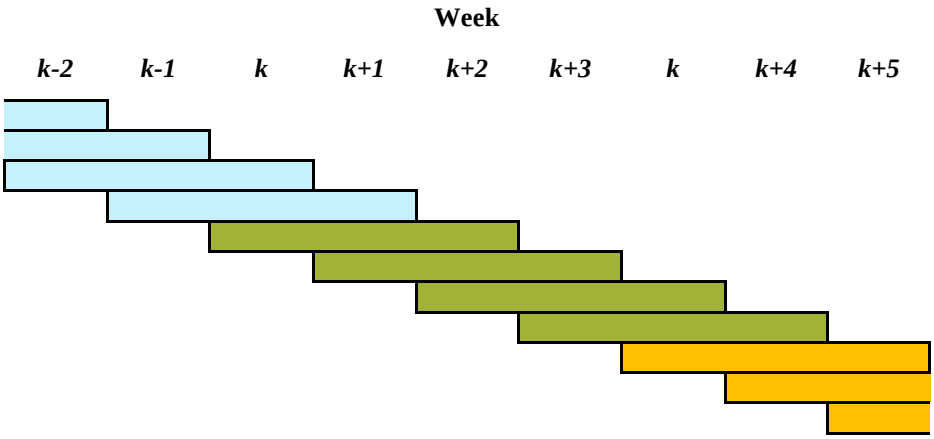


This arrangement of the collection work has the effect that different rounds in an LFS month have different time of field work and also different inflow rate. The first round of the month, for example, have a significantly longer time of field work than the last round, which usually leads to a higher response rate for the first round. The arrangement also places too much focus on the concluding days of the data collection efforts for the LFS month. During the last days, when all four rounds has to be completed at the same

time, the result of the data collection is reconciled and it is not seldom concluded that additional interviews resources and other measures are needed to increase response rates, measures that should have been taken much earlier in the field work in order to have effect. Also, the first rounds in the subsequent LFS-month often suffer from the extra efforts made to "save" the current LFS month at the end of the data collection.

The project recommends instead an arrangement where rounds are completed gradually. This will have the effect that fewer rounds take place simultaneously, which should facilitate the allocation of resources. Moreover, a strategy according to Figure 3 where each round is followed up directly after it ends mean that problems in the data collection can be identified and corrected in time.

Figure 3



References

Kreuter F. (2013) *Improving Surveys with Paradata.* Wileys and Sons. Inc. Hoboken, New Jersey.

SCB (2015). *Förslag till indikatorer och övergripande strategi för datainsamlingen i AKU.* Internal project.