

Data collection, monitoring, data cleaning (first group)

Data collection, monitoring, data cleaning are processes that are done before estimation. This kind of field is getting more and more in view. We can divide this field in two processes:

- a) Before the end of data collection (monitoring, interview type, ..., );
- b) After the end of data collection (nonresponse, data cleaning, editing imputation).

Problems:

- a) **Interview type.** Two types of interview nowadays are used very often (via internet and telephone, via internet and interview and others). How to measure nonresponse (model mode effect in nonresponse) or how to estimate parameters when two types of interview (mixed mode) is used were presenting and in this workshop. But what to do if three types of interview are used?
- b) **Nonresponse.** The main problems with nonresponse occur if we have high noncontact rate or large amount of refusals. The other problem is when it is needed to send numbers of nonresponse by different types for Eurostat – there are different definition for nonresponse types or not all type of nonresponse are counted in different countries.
- c) **Editing.** Like Pauli Ollila said: “Editing (not imputation) is very grey and great area. It is a challenge for methodologist to make a thing brighter.”

Some solutions (proposals):

- a) **Imputation** – if you need to do imputation, but the common imputation methods are not working - use robust imputation (you can contact Nicklas Pettersson for consultation).
- b) **Editing** – Latvia bank does a survey for enterprisers, where about 80 % fill a questionnaire via internet. For these enterprisers the errors of the data are checked during the filling. Also, because most of the enterprisers are filling such kind of questionnaire from time to time, so the data can be compared with the earliest one. You can contact Janis Lapins for more information.
- c) **Data collection** – Central Statistical Bureau of Latvia has new program for data collection. In this program all types of interview (via internet, telephone and face to face interview) are incorporative. Using this program specialist can check different characteristics (response rate, number of questioned units by interviewers and other) during the data collection period. You can contact Martins Liberts for more information.
- d) **Nonresponse** – to reduce the difference between responses and nonresponses response balanced could be used.