

WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

A. Data processing: Processing of data with panel design and multi-mode data collection. Quality Controls and checks of data

A1– Data processing during and after the interview

Cor Kragt – The Netherlands

Data processing during and after the interview

Cor Kragt (Statistics Netherlands)

Workshop LFS methodology, May 2012

Why this subject?

A redesign of the social statistics in the Netherlands stimulated us in 2010 to renew the process for the data processing of the Dutch Labour Force Survey. There were several reasons to create a whole new process. For instance, there were several parts of the software with no function anymore. And also because of the many dependences between the different production lines, which required a good administration on in which order the run the jobs. This is always prone to errors. We therefore aimed at a decrease of dependences between the used production lines. We also needed to include a new weighting model which makes use of more register data than the one we used before.

All this together made us decide to make an new process. In this process we have to deal with some aspects which are typical for a panel survey. We also have to deal with data which are gathered by three different modes of interviewing: face-to-face (CAPI), telephone (CATI) and internet (CAWI).

The purpose of this presentation is to show some aspects of the new data process in the Netherlands as to share our experience.

Contents

Because of a redesign of the social statistics in the Netherlands the Dutch Labour Force Survey is since 2010 in a process of change. Not only the questions and the routing trough the questionnaire has changed. There also has been a change in the approach of households.

Since 1999, when we introduced the panel design in the Netherlands, we used CAPI interviewing in the first wave and CATI interviewing in all of the four consecutive waves. In 2010 the first wave became multi-mode, that is, CAPI as well as CATI is used to collect the data. This year also internet will be used as a mode to collect the data of the first wave.

In 2010 we made a whole new process for the data processing of the Dutch LFS. This year we have to make some adjustments to this process to enable us to handle also the data collected by internet.

The data processing takes places in several stages. The first kind of data processing already takes place during the interview of a household. In the consecutive waves the answers concerning questions on for instance their jobs, working hours, searching for work, occupation, enterprise and education given in the previous wave, are submitted to the respondent. For example, when the respondent confirms that the situation concerning his job has not changed since the last wave, the information about this job is copied forward. So the respondent doesn't have to answer questions about his job again.

However, most of the data processing takes place monthly after finishing the fieldwork period.

In a totally automated system, mostly written in Manipula, all interview records of one month are gathered in a file and different kinds of manipulations are

successively performed on the records of the file: conversion to a unique data model, derivation of variables, matching information from the previous wave, matching the codes for occupation, enterprise and education, weighting and tabulation.

In the presentation we will have a closer look at three aspects of the data processing:

- the conversion to a unique data model has advantages in a process where data is handled which is gathered by different modes with their own data models.
- the matching of information from the previous wave is necessary to complete the records of the consecutive waves, because some information is only determined in the first wave.
- the matching of the codes for occupation, enterprise and education is necessary because these codes are determined in a separate process.

Subjects of discussion

- Because in a panel design the data of one month partly depends on the data of the previous months it is difficult to add new variables or carry out corrections on variables. What is a good way to do this?