The General Population Survey

Throughout this book theoretical concepts are applied on a real survey data set. This data set is based on a Dutch survey that has been carried by Statistics Netherlands. To avoid the disclosure of sensitive individual information, the data set has been anonymized. It is called the General Population Survey (GPS).

The fieldwork of the GPS covered a period of two months. In the first month, selected persons where approached with CAPI (*Computer Assisted Personal Interviewing*). For persons that could not be contacted or refused and who had a listed phone number, a second attempt was made in the second month using CATI (*Computer Assisted Telephone Interviewing*). Table 1 contains the fieldwork results. Note there a nonresponse category 'unprocessed'. This denotes nonresponse caused by unprocessed cases. Such cases were assigned to interviewers, but were not taken up by interviewers due to capacity problems of illness of the interviewer.

Table 1. The fieldwork results of the GPS

Result	Frequency	Percentage
Sample size	32,019	100.0 %
Response	18,792	58.7 %
First month	14,870	46.4 %
Second month	3,922	12.2 %
Non-response	13,227	41,3 %
Unprocessed cases	2,456	7.7 %
Non-contact	1,847	5.8 %
Refusal	7,890	24.6 %
Not able	1,034	3.2 %

Persons were selected by means of a stratified two-stage sample. In the first stage, municipalities were selected within regions with probabilities proportional to the number of inhabitants. In the second stage, an equal probability sample was drawn in each selected municipality. Sampling frames for the persons were the population registers of the municipalities. The sampling design was such that each person had the same probability of being (a so-called self-weighting sample). The sample of the GPS consisted of 32,019 persons. The number of respondents was 18,792.

Statistics Netherlands has an integrated system of social statistics. This system is called the *Social Statistics Database* (SSD). The SSD contains a wide range of characteristics on each individual in the Netherlands. There are data on demography, geography, income, labor, education, health, and social protection. These data are obtained by combining data from registers and other administrative data sources.

SSD records could be linked to the survey data records using a personal identification numbers. This can be done both for respondents and non-respondents. Thus, demographic variables like sex, age, province of residence, and ethnicity became available for all sampled persons, and also socio-economic variables like employment and various types of social security benefits.

The Netherlands is divided in approximately 420,000 postal code areas. A postal code area contains, on average, 17 addresses. These areas are homogeneous with respect to social and economic

characteristics of its inhabitants. Using information from the population register, Statistics Netherlands has computed some demographic characteristics for these postal code areas. Since postal codes are included in the survey data file for both respondents and non-respondents, these characteristics can be linked to the survey data file. Examples of such variables are degree of urbanization, town size, percentage of people with a foreign background (non-natives). From another source also the average house value was included.

During the fieldwork period, interviewers have kept record of all contact attempts. For each attempt its contact result was recorded (contact, or not). In case contact was established, the result of the co-operation request was recorded (response or non-response, and in case of non-response the reason of non-response). Also other information was included, like the mode of the fieldwork attempt (CAPI or CATI). All this fieldwork information is available for analysis.

Table 2 contains an overview of all variable in the survey data file of the GPS. The values of the target variables are only available for the respondents. The auxiliary variables are available for both respondents and non-respondents.

Table 2. The variables in the GPS survey data file

Variable	Description	Cats	Type	Source
PC	PC in household	2	Target variable	Survey
Move	Wants to move within a year	6	Target variable	Survey
Health	General health condition	5	Target variable	Survey
Newspap	Has newspaper subscription	2	Target variable	Survey
Club	Is active in a club	2	Target variable	Survey
Politics	Is interested politics	4	Target variable	Survey
Employed	Employment situation	3	Target variable	Survey
Educat	Level of education	8	Target variable	Survey
Ownhouse	Owns house	2	Target variable	Survey
Religion	Religious denomination	5	Target variable	Survey
Gender	Gender	2	Auxiliary variable	SSD
Marstat	Marital status	4	Auxiliary variable	SSD
Married	Is married	2	Auxiliary variable	SSD
Age3	Age in 3 age groups	3	Auxiliary variable	SSD
Age13	Age in 13 age groups	13	Auxiliary variable	SSD
Nonnativ	Is non-native	2	Auxiliary variable	SSD
Ethnic	Type of non-native	5	Auxiliary variable	SSD
HHSize	Size of the household	5	Auxiliary variable	SSD
HHType	Type of household	5	Auxiliary variable	SSD
Children	Children in household	2	Auxiliary variable	SSD
Phone	Has listed phone number	2	Auxiliary variable	SSD
Hasjob	Has a job	2	Auxiliary variable	SSD
Employed	Employment situation	3	Auxiliary variable	SSD
Socall	Has social allowance	2	Auxiliary variable	SSD
Disaball	Has disability allowance	2	Auxiliary variable	SSD
Unempall	Has unemployment allowance	2	Auxiliary variable	SSD
Allowan	Has an allowance	2	Auxiliary variable	SSD
Region	Region of the country	5	Auxiliary variable	SSD
Urban	Degree of urbanization	5	Auxiliary variable	SSD
Houseval	Average house value in neighborhood	12	Auxiliary variable	SSD
Pnonnat1	Percentage non-natives in neighborhood	8	Auxiliary variable	SSD
Pnonnat2	Percentage non-western non-natives in neighborhood	7	Auxiliary variable	SSD
Resp1	Response in first month of fieldwork	2	Fieldwork variable	SSD
Result	Final result of the fieldwork attempt	5	Fieldwork variable	SSD
Response	Response	2	Fieldwork variable	SSD

There were 150 cases in the initial sample that did not belong to the target population of the GPS. These cases are not contained in GPS survey data file. The amount of over-coverage is needed to be able to compute the response rate of the GPS.