

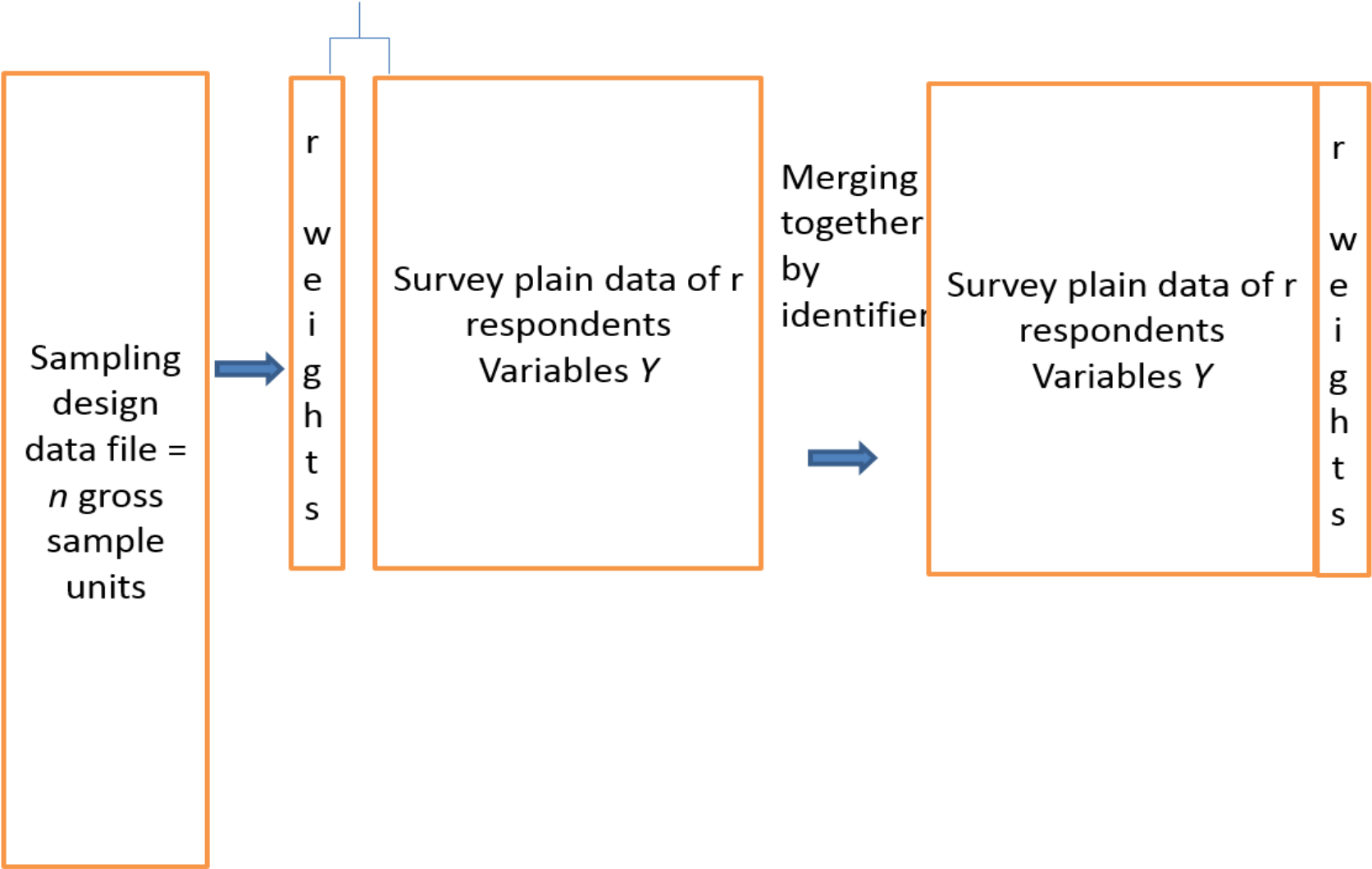
# Chapter 5

## Sampling Design Data File



## **Sampling design data file**

The term 'sampling design (data) file' that is not commonly used in survey sampling literature. The methodology behind this term is used, but implicitly. Its explicit determination facilitates many things in survey practice and also gives a clear target for the two big parts of a survey, that is, sampling and fieldwork. The sampling design file consists of all the gross sample units and its variables include those that give opportunity to create sampling weights and to analyse the survey quality. The file is possible to complete after the fieldwork. Its most important characteristics, including sampling design variables and weights, will be finally merged together with the real survey variables at respondent level, and then the survey analysis is ready to start. Scheme 5.1 elucidates this situation.



The following list consists of the variables that should or could be included in the sampling design file, with good meta data:

- (i) Inclusion probabilities of each stage
- (ii) Other variables directly relating to the sampling design (*PSU* that can be a cluster or an individual, explicit stratum, implicit stratum)
- (iii) Outcome of the survey fieldwork (respondent, ineligible, non-respondent)
- (iv) Macro auxiliary variables, statistics for the small area level such as PSU, stratum, grid square: population by gender or by age group or by education degree)
- (v) Macro auxiliary variables, as rates for small areas as in point (iv): % divorced, % under crowding, % 2 or more cars, % owner occupation, % unemployed, % long term unemployed, % social renting, % highly educated

<b>Variable</b>	<b>Label</b>	<b>Categories</b>
IDNO	Respondent's identification number	
PROB1	Inclusion probability first sampling stage (small area)	
PROB2	inclusion probability second sampling stage (dwelling)	
PROB3	Inclusion probability third sampling stage (15+ years old individual)	
age	Age of sampled individual	
citizen	Domestic vs not citizenship	2
common_law_marriage	Common law marriage	2
education	Education level	6
gndr	Gender of selected individual	2
hinctnta	Household's total net income, all sources	10
maritalb	Legal marital status, post coded	
member15	Members below 15	
members15Plus	Members 15 years old or more	
outcome	Fieldwork outcome	3
PSU	Primary sampling unit	
stratum	Explicit strata	8
agegroup	15, 30, 45, 60, 75+	5

## 20 observations from the test data file

	Primary sampling unit	Explicit stratum	Respondent identification number	Inclusion probability of 1st stage	Inclusion probability of 2nd stage	Inclusion probability of 3rd stage, individual	Citizenship target person (taken from register)	Education level	Members below 15	outcome
1	1	1	370	0.001527951915	0.352941176471	0.250000000000	1	2	0	2
2	1	1	467	0.001527951915	0.352941176471	0.250000000000	2	3	0	2
3	1	1	567	0.001527951915	0.352941176471	1.000000000000	1	3	0	2
4	1	1	1530	0.001527951915	0.352941176471	0.500000000000	1	3	2	1
5	1	1	1745	0.001527951915	0.352941176471	0.333333333333	1	6	0	2
6	1	1	2501	0.001527951915	0.352941176471	0.500000000000	1	4	2	2
7	1	1	2569	0.001527951915	0.352941176471	0.500000000000	2	7	2	1
8	1	1	3135	0.001527951915	0.352941176471	0.500000000000	1	5	1	1
9	1	1	3451	0.001527951915	0.352941176471	0.500000000000	1	3	2	1
10	1	1	3673	0.001527951915	0.352941176471	0.333333333333	1	5	1	2
11	1	1	3916	0.001527951915	0.352941176471	0.250000000000	1	3	0	1
12	1	1	4607	0.001527951915	0.352941176471	1.000000000000	1	3	0	2
13	1	1	5016	0.001527951915	0.352941176471	0.500000000000	1	3	0	3
14	1	1	6246	0.001527951915	0.352941176471	1.000000000000	1	7	0	1
15	1	1	6247	0.001527951915	0.352941176471	0.500000000000	1	3	0	1
16	1	1	7514	0.001527951915	0.352941176471	0.500000000000	1	7	0	1
17	1	1	8319	0.001527951915	0.352941176471	0.500000000000	1	2	0	2
18	1	1	8444	0.001527951915	0.352941176471	0.500000000000	1	7	0	1
19	1	1	8525	0.001527951915	0.352941176471	1.000000000000	1	2	0	2
20	1	1	8845	0.001527951915	0.352941176471	0.333333333333	1	3	0	1