Topics in Social Statistics
University of Helsinki
Part I _A

Seppo Laaksonen

Survey Concepts

Content of Part I

What is survey?
Key concepts in surveys
From Survey Data Collection to Cleaned Survey Data

Through
- Designing the survey
- Designing the questionnaire
- Designing the sample(s)
- Data collection with alternative single and mixed modes
- Data entry
- Editing the raw data
- Imputing the data
- Weighting the data
- Adding other features into the data file

Part 3 extends some to Part 1, especially in analysing cleaned data. Part 2 is focused on measurement questions in data collection and analysis.

What is survey?

It is a demanding question to answer.

Wikipedia:
Statistical surveys are used to collect quantitative information about items in a population. Surveys of human populations and institutions are common in political polling and government, health, social science and marketing research. A survey may focus on opinions or factual information depending on its purpose, and many surveys involve administering questions to individuals. When the questions are administered by a researcher, the survey is called a structured interview or a researcher-administered survey. When the questions are administered by the respondent, the survey is referred to as a questionnaire or a self-administered survey.

What is survey?

It is a demanding question to answer.
For me:
Survey is a series of tasks that finally results a statistical file of statistical units and their characteristics (variables). These units may be:
- Individual people
- Households
- Families
- Enterprises
- Plants (Local units of enterprises)
- Local-kind-of-activity units of enterprises
- Villages
- Municipalities
- Other areas
- Societies

Such a data file may cover basically the whole population (including register) or it can be based on a sample (= survey sampling, survey statistics).

My recent definition

Source: Encyclopedia of Behavioral Medicine, 2012. Springer
The encyclopedia cover can be viewed at
http://www.springer.com/medicine/book/978-1-4419-1004-2?changeHeader.

Survey is a methodology and a practical tool used to collect, handle, and analyze in a systematic way information from individuals. These individuals or micro units can be of various types, such as people, households, hospitals, schools, businesses, or other corporations. The units can be simultaneously available from two or more levels such from households and their members. Information in surveys may be concerned various topics such as people's personal characteristics, their behaviour, health, salary, attitudes and opinions, incomes, poverty and housing environments, or characteristics and performance of businesses. Survey research is unavoidably inter-disciplinary, although the role of statistics is most influential since the data for surveys is constructed in a quantitative form. Correspondingly, many survey methods are special statistical applications. However, surveys exploit substantially many other sciences such as informatics, mathematics, cognitive psychology, and theory of subject-matter sciences of each survey topic.

Comments?

Key concepts in surveys

Next pages concentrate on
- Populations in surveys
- Cross-sectional vs. longitudinal surveys
- Other basic concepts such as Meta data and para data

After that I go to
- Sampling design
- Missingnesses and other deficiencies
- Editing
- Imputation
- Sampling and other weights

Populations in surveys 1

In statistics population is a key concept determined by Adolphe Quetelet in 1820's. This is not just one in surveys where I need even five populations:

1. *Population of interest* is the population that a user would like to get or estimate ideally but it is not possible always to completely reach and hence she/he determines
2. *Target population* which is such a population that is realistic. Naturally, this population should be exactly determined including its reference period (a point of time or a time period).

The target population of the ESS e.g. "Persons 15 years or older who are resident within private households in the country in the first of November."
Correspondingly to the EFSS (European Finnish Security Survey): 15-74 years old non-Swedish speaking residents in Finland 1st of October 2009.
Give your own example.

Populations in surveys 2

In order to get the target population you need

3. *Frame population and the frame* from which the statistical units for the survey can be found. Usually, the frame is not exactly from the same period as the target population (delay in Finnish population surveys is rather short i.e. 1-5 months, but for enterprise surveys much more, even some years).

The frame is not always at element level available as in the case of Finnish population register based surveys. Instead, the frame population can be as follows:

Stage 1: List of the electoral sections (e.g. in a certain country their number is 12,313 and they cover the whole territory of the country).
Stage 2: Lists of all households' addresses of the at the first stage selected units.
Stage 3: One or more members of the selected household/address

There are here thus two frames, but it is possible that this number can be even four such municipalities, blocks or villages or census districts, addresses, people at certain ages.

Populations in surveys 3

Due to the delay in the frame,

4. *Updated frame population* is useful for estimating the results better. Usually, the initial frame population has been used for estimation too. This may lead to biased estimates. Fortunately, this bias is not severe in most human surveys. At contrast, old frames can lead to dramatic biases in business surveys, if this is due to large businesses.

After the data collection or fieldwork we are able to determine

5. *Study population or survey population.*

It is ideal if this fifth population corresponds to our target population or even the population of interest. But if not, our estimates are somewhat biased.

Populations in surveys 4

The units of the target population are equal to those of the study population but the units of the frame population(s) can be essentially different.

The ESS survey designs vary a lot from one country to the next. There are such countries where all the units are equal = individuals 15+ (Finland, Sweden, …) but many countries have several units (small areas, addresses, households, 15+ years individuals, …).

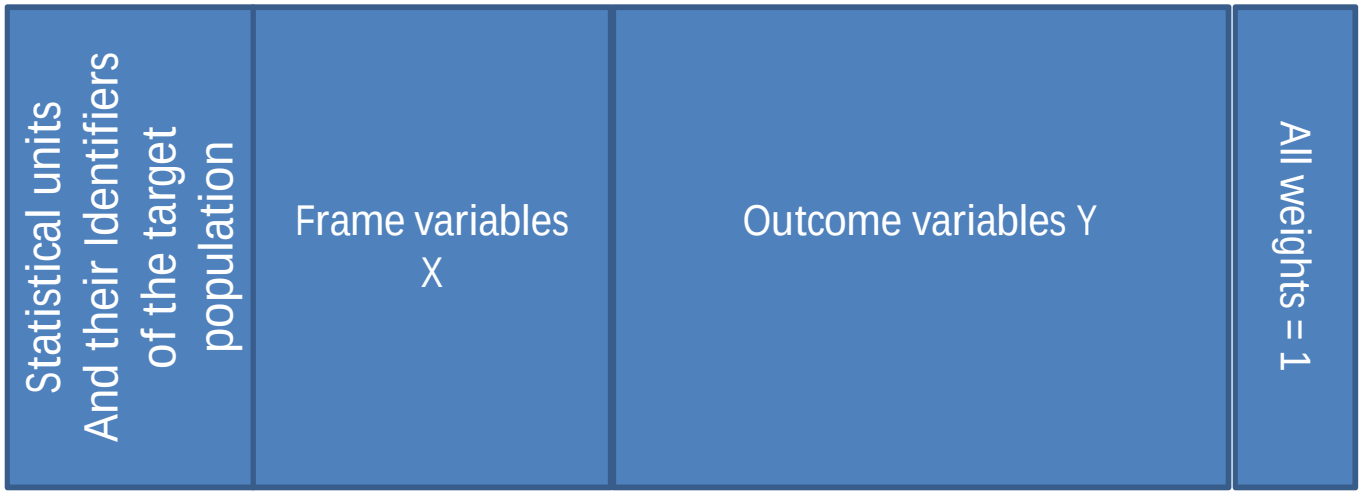
PISA and other student surveys use typically two units:
- 'PISA' Schools (or school classes)
and
- Students themselves

Give other examples.

The next two pages illustrate missingnesses as well as some other crucial concepts in surveys. The first is the case of a cross-sectional survey.

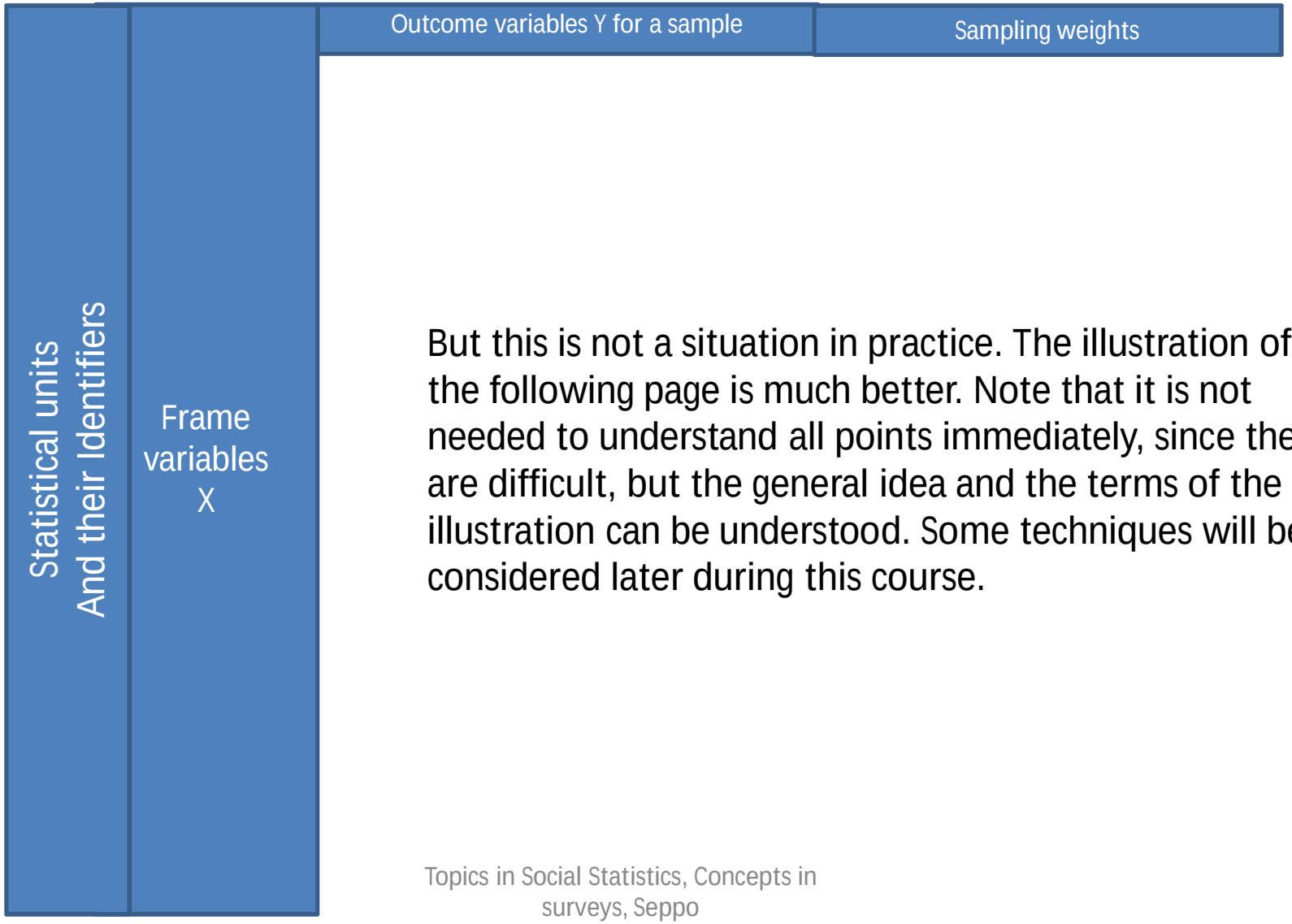
Micro data for the whole target population

Now I focus on micro data. I start with a cross-sectional case. If the whole target population has been examined, and any missingnesses occur, it is simple as the following scheme illustrates.



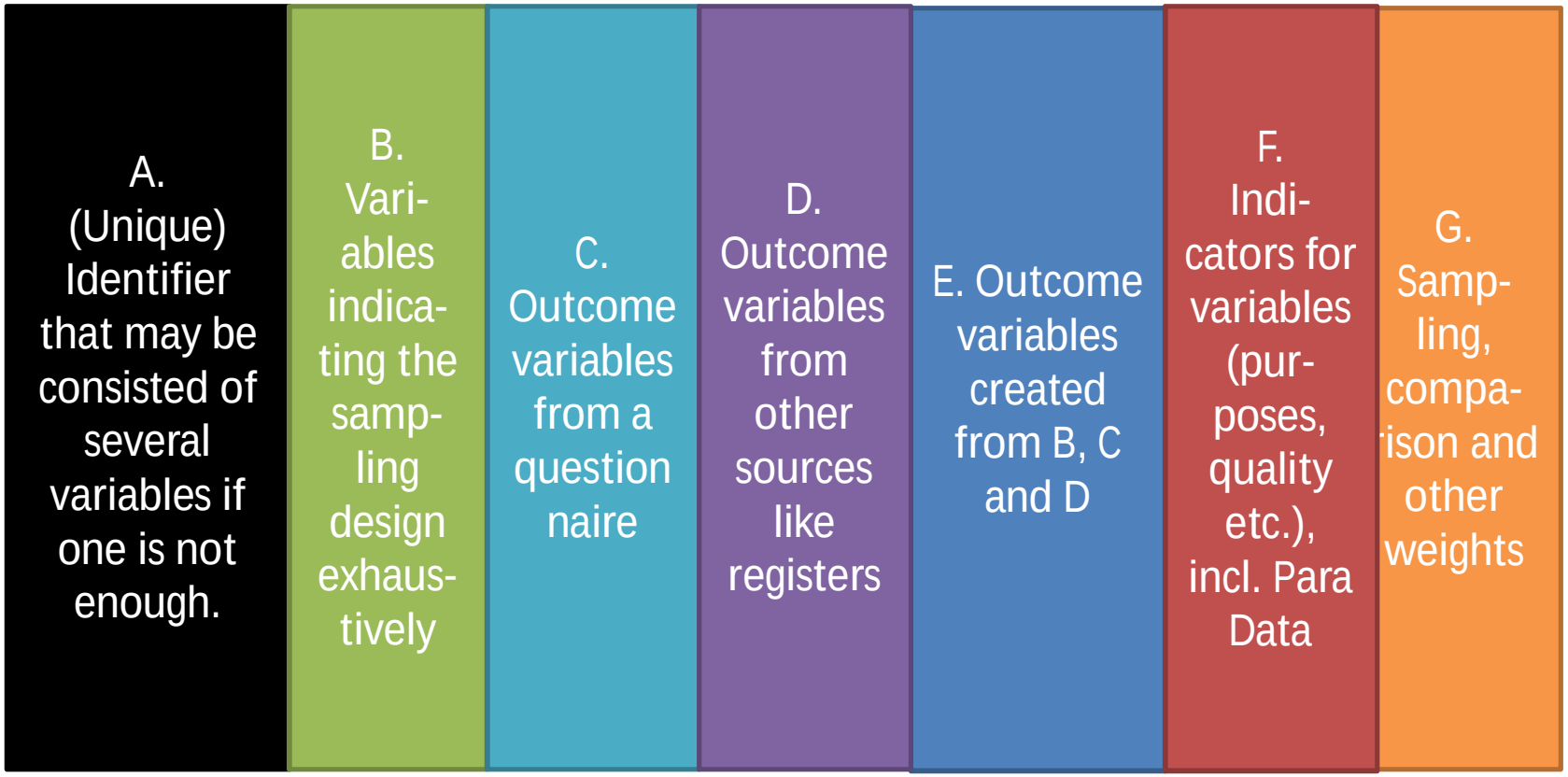
Micro data without missingness

But most surveys are based on a sample. Below is a simple illustration for this case without non-response :

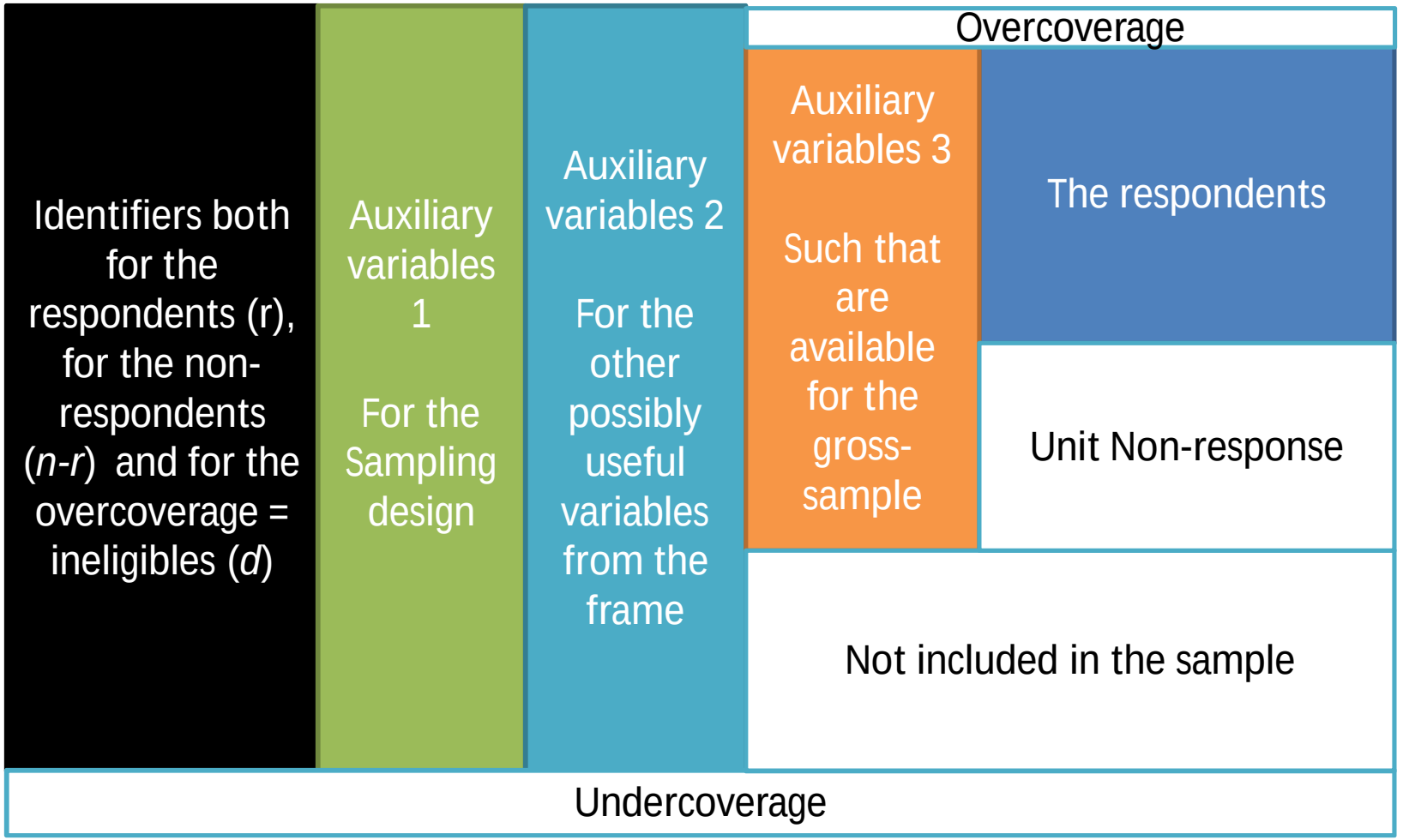


But this is not a situation in practice. The illustration of the following page is much better. Note that it is not needed to understand all points immediately, since they are difficult, but the general idea and the terms of the illustration can be understood. Some techniques will be considered later during this course.

General structure of a micro level cross-sectional survey data file that consists of *r* respondents (rows of the matrix).



General scheme of the cross-sectional micro file in which the previous is one box



Micro data and Missingness

Examples of the terms in social surveys:

Overcoverage (in-eligibles): died, emigrants, errors in the frame

Some of them can be observed during the fieldwork, not all

Undercoverage: new born, new immigrants, errors in the frame

Updated frame helps to find them

Unit non-response: not contacted, disable to participate, refusals, ...

Sampling weights are of two types:

- Their average is = 1 and hence their sum = the number of the respondents
- Their sum = the number of the target population units (households or individuals, etc.) and the average is how much one unit represents in the estimation.

A real survey file is not such as the scheme of the below page, except in some special cases like methodological experiments. There are two real files:

- Sampling file that covers the gross sample units and auxiliary variables. I have a special (good) example in another document of the course website. From this file we usually create the sampling weights and other sampling design variables and merged these into 
- The file of the respondents that is just used in analysis. In the end of this section, I give examples of such survey files and of course we will later use these in our analysis.

Comment on variables:

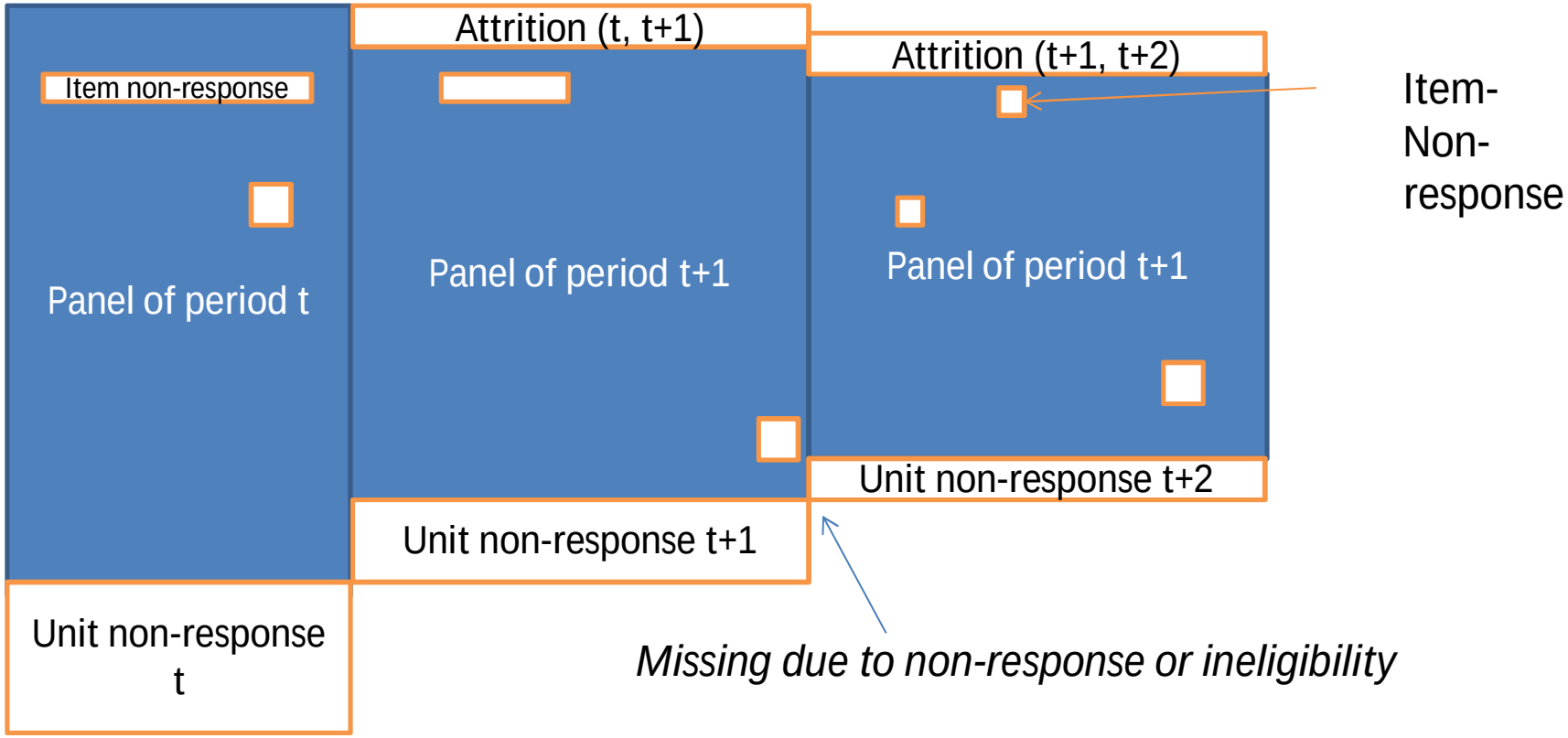
As noticed, there are X and Y variables and they have a special role. However, a X variable can be used as a Y variable as well in the analysis but in this case, their values are only for the respondents.

Moreover, Y variables can be of different kinds:

- Initial survey questionnaire variables and exactly in the same form as in the questionnaire.
- But these initial Y variables can be transformed into another form (a new scale) as well in order to facilitate analysis.
- Summary variables from another source like student (PISA) exams, clinical examinations, ...
- Aggregated information e.g. from a living area characteristics (the same value for all living in this area).
- What else?

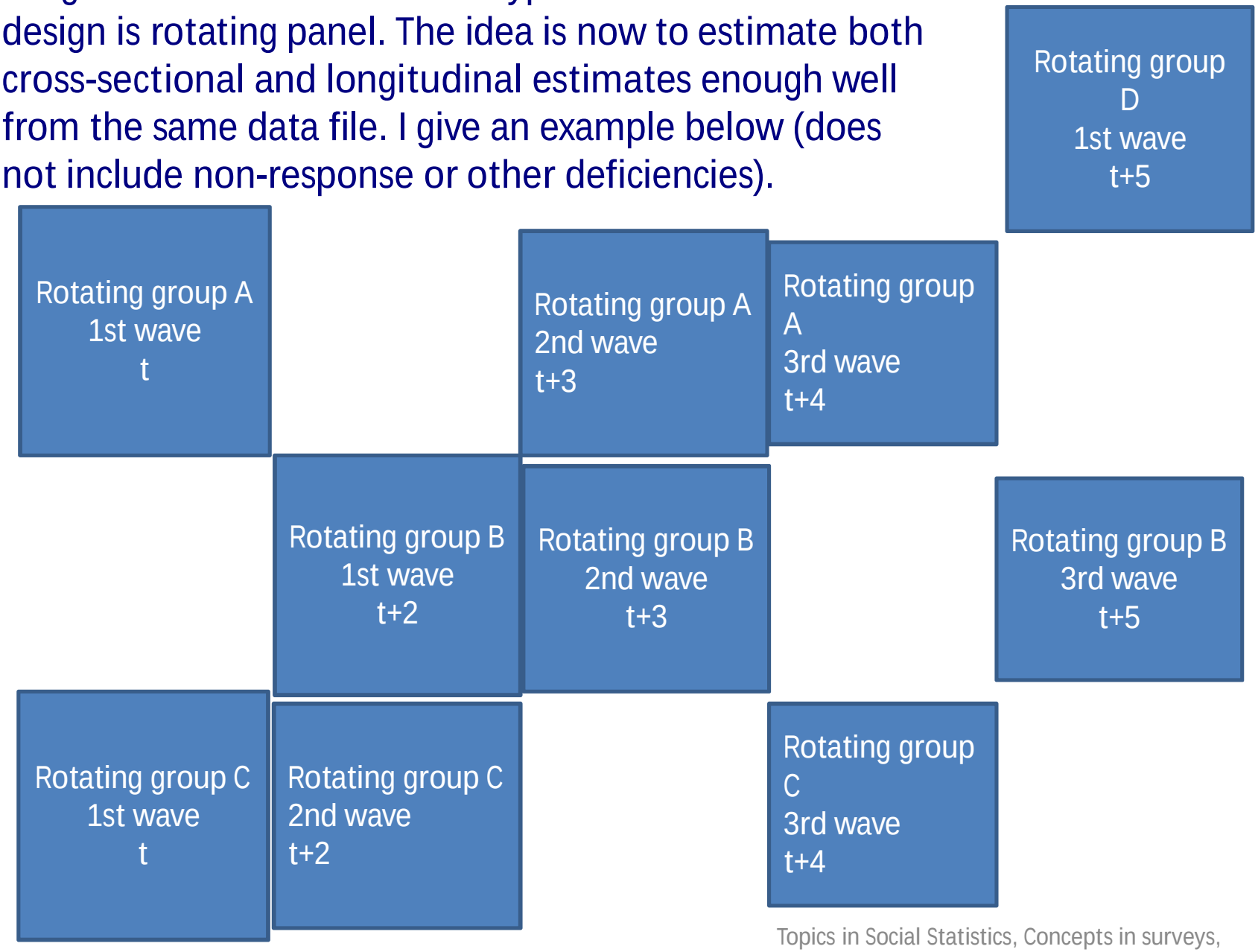
Micro data and Missingness

Cohort type of panel (longitudinal) example

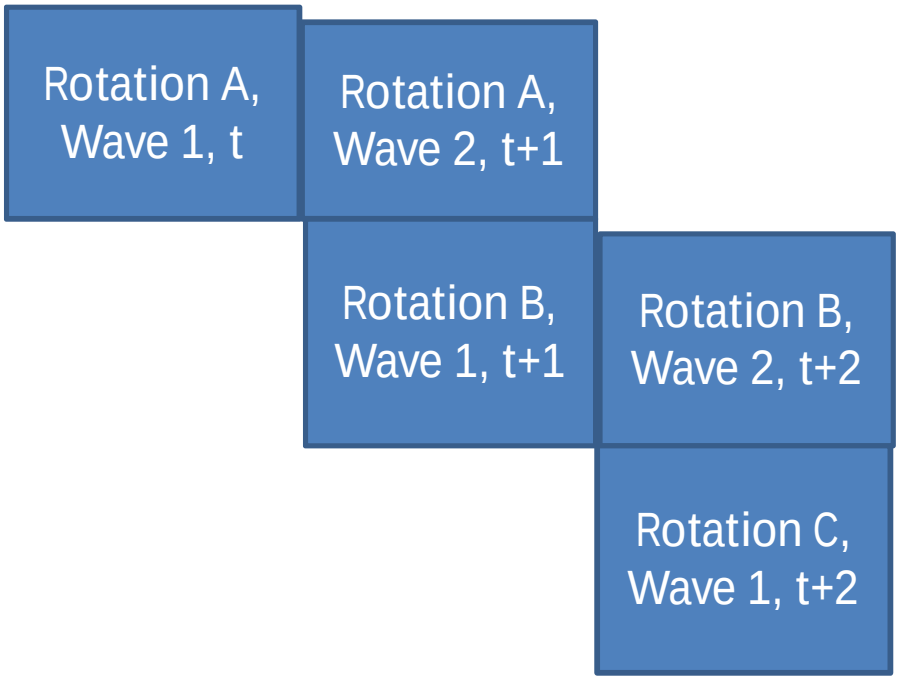


Rotating panel

Longitudinal can be of various types. One common design is rotating panel. The idea is now to estimate both cross-sectional and longitudinal estimates enough well from the same data file. I give an example below (does not include non-response or other deficiencies).



I have applied 20 years ago a rotating panel design for the Statistics Finland Income survey. This is maybe the simplest possible rotating design, since the panel covers only two years.



A more complex rotation has been used e.g. in Labour force surveys of all European countries. The reason is that they wish to e.g. estimate both unemployed rates and changes in these rates in their longitudinal meanings. Most income surveys have the same purpose, e.g. how permanent is poverty or riches?

Survey micro data that is hard to understand. Why?

	ST01Q01	ST08Q01	ST09Q01	ST10Q01	ST13Q01	ST20Q04	ST21Q04	ST23Q01	ST25Q01	ST33Q01	AGE	ATTCOMP	ATSCHL	PV1MATH	PV1READ	PV1SCIE	W_FSTUWT	STRATUM
1	9	1	3221	1	6111	1	3	2	5	2	15,25	,86	-,44	522,44	462,05	457,42	14,16	34802
2	10	1	4222	1	3422	1	3	5	1	2	15,92	-1,42	-1,13	591,38	611,91	582,56	29,15	34802
3	9	1	3150	1	7136	1	1	5	5	1	16,08	,86	1,09	543,16	599,67	600,28	20,30	34802
4	10	1	6112	1	9999	1	2	4	5	2	15,83	-,11	,51	569,64	571,08	557,39	25,16	34802
5	9	1	8263	2	8320	1	2	2	3	2	15,75	-,11	-,44	398,67	469,95	469,45	17,51	34803
6	9	1	7122	3	8324	1	2	1	1	1	15,42	,86	-1,84	417,36	315,51	395,79	27,51	34804
7	10	1	9505	1	9505	1	3	1	3	2	16,00	-,11	-,44	501,49	522,46	501,44	16,87	34803
8	10	1	2359	1	2310	1	3	4	2	2	16,00	,86	-,44	490,19	499,50	503,95	16,56	34802
9	8	1	2141	.	9504	1	1	3	3	3	15,25	-,11	-,44	488,25	502,60	420,96	48,23	34801
10	9	1	2140	1	2144	1	2	3	2	1	15,67	-,11	2,01	606,64	603,31	614,92	21,24	34802
11	9	1	6113	2	7124	1	1	4	5	1	15,58	-,75	1,09	470,25	569,65	519,25	18,16	34802
12	9	1	5220	1	2224	1	3	2	2	2	15,67	-,75	,02	467,99	452,59	487,17	26,61	34802
13	9	1	9504	3	7422	1	2	2	2	2	15,33	,86	-,44	478,90	498,31	497,71	17,65	34803
14	9	1	9503	1	7311	1	3	1	5	2	15,58	,86	.	490,12	491,16	495,84	23,09	34803
15	9	1	2441	1	2443	1	1	4	2	2	15,25	-,75	-,82	596,05	605,16	634,78	18,78	34802
16	10	1	2222	1	2221	1	3	1	4	2	16,00	-,11	,02	638,11	590,24	578,83	19,84	34802
17	10	1	3431	1	3150	1	2	4	3	3	16,08	,86	-1,13	456,23	505,30	502,46	16,40	34803
18	9	1	7220	3	7245	1	1	2	5	1	15,67	-2,44	1,09	373,04	398,93	352,98	22,03	34803
19	9	2	8263	3	9504	2	2	2	5	2	15,50	,86	-,44	518,23	490,84	534,07	18,37	34803
20	9	1	8263	2	7224	1	2	2	5	1	16,17	,86	2,01	348,35	426,81	449,22	23,71	34804
21	9	1	5220	2	7100	1	2	2	5	1	15,92	,86	2,01	388,54	361,43	415,46	17,70	34804
22	10	1	9999	2	9999	1	2	5	5	2	15,92	-1,42	,51	496,89	603,88	563,73	14,23	34803
23	9	1	9132	3	5143	1	1	2	5	1	16,17	,86	,02	458,72	464,86	463,02	19,60	34804
24	10	1	8240	3	9503	1	2	2	4	2	16,17	-1,42	-,44	377,25	488,46	452,95	40,67	34803
25	9	1	9141	1	8320	1	1	2	4	1	15,50	-,11	1,09	446,88	440,95	417,61	21,06	34803
26	10	1	2332	1	7137	1	2	2	3	2	16,17	-1,42	-,44	589,51	537,89	550,67	14,26	34802
27	9	1	2321	1	1310	1	4	2	5	2	15,42	-,75	-,82	611,32	602,69	588,81	20,51	34802
28	9	1	8263	3	7224	1	2	3	5	3	15,50	,86	-,82	543,55	513,60	554,59	20,97	34803
29	10	1	2145	1	2141	1	2	2	5	2	16,17	-,75	-1,39	517,22	602,14	585,64	19,62	34802
30	7	1	8290	3	6141	2	1	2	5	1	16,17	-,11	2,01	405,68	405,36	392,33	52,25	34801
31	10	1	3221	1	2221	2	3	1	2	2	16,17	-,11	,02	522,44	489,31	484,46	18,34	34803
32	9	1	1233	1	5169	1	2	2	5	2	15,58	,86	,51	493,31	464,86	496,77	21,31	34803
33	10	1	7212	2	5112	1	2	1	5	1	16,00	-,11	2,01	448,68	537,79	473,74	14,16	34802
34	9	1	8263	2	5122	1	2	2	4	2	15,50	-,11	-,44	492,14	555,77	511,69	21,31	34803
35	10	1	9132	2	1210	1	2	5	3	1	16,17	-,11	2,01	489,02	534,37	492,11	20,59	34803
36	9	1	3221	1	8324	1	3	2	5	2	15,58	,86	,02	488,95	465,26	481,67	22,64	34803
37	9	1	8211	3	9141	1	1	2	1	.	15,42	-,75	,07	406,77	429,59	442,69	21,06	34803

This is easier. Why? This Is SAS. Previous SPSS, but the data are the same. Which?

VIEWTABLE: Work.Hungary

	SCHOOLID	StIDStd	ST01Q	ST08Q0	ST09Q01	ST10Q01	ST13Q01	ST2	ST21Q	ST23Q01	ST25Q01	ST33Q01	AGE	ATTCOM	ATSCH	PV1MAT	PV1RE	PV1SCI	W_FSTU	STRATUM
1	00085	02135	10	Yes	Nursing Associate Professionals [Incl. Trainee Nurses]	<ISCED level 3A>	Nursing Associate Professionals [Incl. Trainee Nurses]	Yes	One	30 minutes or less a day	Several times a month	Disagree	15.83	-0.7516	-0.821	515.12	442.33	453.6	19.6939	HUN: vocational secondary school
2	00130	03242	9	Yes	Shop Salespersons & Demonstrator	<ISCED level 3B, 3C>	Plant & Machine Operators & Assemblers	Yes	None	I don't read for enjoyment	About once a month	Strongly disagree	15.33	0.8605	2.0085	412.22	418.47	465.07	17.6506	HUN: vocational secondary school
3	00010	00258	9	Yes	Shop Salespersons & Demonstrator	<ISCED level 3A>	Building Etc Electricians	Yes	One	1 to 2 hours a day	Several times a week	Strongly disagree	15.67	-0.108	1.0878	529.14	598.8	593.47	24.3589	HUN: grammar school
4	00054	01315	9	Yes	Nursing Associate Professionals [Incl. Trainee Nurses]	<ISCED level 3A>	Carpenters & Joiners	Yes	Two	30 minutes or less a day	A few times a year	Disagree	15.58	-0.108	0.0185	548.54	540.54	550.86	22.0277	HUN: grammar school
5	00096	02404	10	Yes	Missing	<ISCED level 2>	Travel Attendants & Travel Stewards	Yes	One	30 minutes or less a day	Several times a month	Agree	16.08	M	-1.1303	348.04	382.85	409.77	17.1453	HUN: vocational school
6	00069	01723	9	Yes	Waiters, Waitresses & Bartenders	<ISCED level 3A>	Protective Services Workers [Incl. Bodyguard, Coastguard]	Yes	None	Between 30 and 60 minutes	Several times a week	Disagree	15.5	0.8605	-0.821	522.6	512.4	553.47	20.6607	HUN: vocational secondary school
7	00154	03878	9	Yes	Lawyers	<ISCED level 3A>	Lawyers	Yes	Two	Between 30 and 60 minutes	Several times a week	Strongly disagree	15.58	-0.7516	0.5091	671.22	668.89	660.7	21.2408	HUN: grammar school
8	00115	02854	9	Yes	Other Associate Professionals	<ISCED level 3A>	Heavy Truck & Lorry Drivers	Yes	One	30 minutes or less a day	Several times a week	Strongly disagree	15.42	0.8605	0.5091	524.08	567.74	550.58	24.0007	HUN: grammar school
9	00071	01780	10	Yes	Other Teaching Professionals Nec	<ISCED level 3A>	Higher Education Teaching Professionals [Incl. Univ. Professor]	Yes	Two	1 to 2 hours a day	A few times a year	Disagree	16	0.8605	-0.4352	490.19	499.5	503.95	16.5571	HUN: grammar school
10	00178	04458	10	Yes	Primary Education Teaching Professionals	<ISCED level 3A>	Primary Education Teaching Professionals	Yes	One	More than 2 hours a day	A few times a year	Strongly disagree	16	-0.108	0.5091	647.93	616.14	633.57	13.9604	HUN: grammar school
11	00065	01621	9	Yes	Primary Education Teaching Professionals	<ISCED level 3A>	[Small Enterprise] General Managers Wholesale & Retail Trade	Yes	Two	30 minutes or less a day	Several times a week	Disagree	15.83	0.8605	0.5091	499.54	535.25	494.63	21.2599	HUN: vocational secondary school
12	00130	03243	9	Yes	Do Not Know	<ISCED level 2>	Cabinet-Makers Etc Workers [Incl. Cartwright, Cooper]	Yes	One	30 minutes or less a day	A few times a year	Disagree	15.33	0.8605	-0.4352	478.9	498.31	497.71	17.6506	HUN: vocational secondary school
13	00010	00261	9	Yes	Finance & Sales Associate Professionals Nec	<ISCED level 3A>	Shop Salespersons & Demonstrators	Yes	Two	30 minutes or less a day	About once a month	Strongly disagree	15.67	0.8605	1.0878	520.57	477.69	509.83	27.1997	HUN: grammar school
14	00063	01562	9	Yes	Precision-Inst Makers & Repairers [Incl. Dental, Watch Maker]	<ISCED level 2>	Painters Etc Workers [Incl. Construction Painter, Paperhanger]	Yes	One	I don't read for enjoyment	A few times a year	Disagree	15.75	-0.108	0.0185	485.05	453	549.09	22.6412	HUN: vocational secondary school
15	00174	04359	11	Yes	Dentists	<ISCED level 3A>	Dentists	Yes	Two	30 minutes or less a day	About once a month	Strongly disagree	15.83	-0.108	2.0085	643.96	642.43	653.25	18.1435	HUN: grammar school
16	00120	03007	9	Yes	Production Dep. Managers Wholesale & Retail Trade	<ISCED level 3A>	Electrical Engineers	Yes	One	Between 30 and 60 minutes	Never or almost never	Disagree	15.25	0.8605	-0.821	509.59	564.64	550.86	27.7043	HUN: grammar school
17	00093	02304	9	Yes	Sewing-Mac Operators	<ISCED level 3B, 3C>	Heavy Truck & Lorry Drivers	Yes	One	Between 30 and 60 minutes	Several times a week	Disagree	15.42	-0.5293	-0.4352	429.44	397.44	441.48	18.1066	HUN: vocational school

Now our data are quite easy to understand due to meta data exploited.

VIEWTABLE: Work.Hungary

	School ID 5-digit	Student ID 5-digit	GRAD	At Home - Mother	Mother Occupation	Mother <Highest Schooling>	Father Occupation	Poss com	How many cars	Reading Enjoyment Time	Like Read - Magazines	Climate - Little for Adult Life	Age of student	Attitude towards computer	Attitude towards school	Plausible value in math	Plausibl value in reading	Plausible value in science	FINAL STUDEN WEIGHT	Original stratum
1	00085	02135	10	Yes	Nursing Associate Professionals [Incl. Trainee Nurses]	<ISCED level 3A>	Nursing Associate Professionals [Incl. Trainee Nurses]	Yes	One	30 minutes or less a day	Several times a month	Disagree	15.83	-0.7516	-0.821	515.12	442.33	453.6	19.6939	HUN: vocational sec
2	00130	03242	9	Yes	Shop Salespersons & Demonstrator	<ISCED level 3B, 3C>	Plant & Machine Operators & Assemblers	Yes	None	I don't read for enjoyment	About once a month	Strongly disagree	15.33	0.8605	2.0085	412.22	418.47	465.07	17.6506	HUN: vocational secondary school
3	00010	00258	9	Yes	Shop Salespersons & Demonstrator	<ISCED level 3A>	Building Etc Electricians	Yes	One	1 to 2 hours a day	Several times a week	Strongly disagree	15.67	-0.108	1.0878	529.14	598.8	593.47	24.3589	HUN: grammar school
4	00054	01315	9	Yes	Nursing Associate Professionals [Incl. Trainee Nurses]	<ISCED level 3A>	Carpenters & Joiners	Yes	Two	30 minutes or less a day	A few times a year	Disagree	15.58	-0.108	0.0185	548.54	540.54	550.86	22.0277	HUN: grammar school
5	00096	02404	10	Yes	Missing	<ISCED level 2>	Travel Attendants & Travel Stewards	Yes	One	30 minutes or less a day	Several times a month	Agree	16.08	M	-1.1303	348.04	382.85	409.77	17.1453	HUN: vocational school
6	00069	01723	9	Yes	Waiters, Waitresses & Bartenders	<ISCED level 3A>	Protective Services Workers [Incl. Bodyguard, Coastguard]	Yes	None	Between 30 and 60 minutes	Several times a week	Disagree	15.5	0.8605	-0.821	522.6	512.4	553.47	20.6607	HUN: vocational secondary school
7	00154	03878	9	Yes	Lawyers	<ISCED level 3A>	Lawyers	Yes	Two	Between 30 and 60 minutes	Several times a week	Strongly disagree	15.58	-0.7516	0.5091	671.22	668.89	660.7	21.2408	HUN: grammar school
8	00115	02854	9	Yes	Other Associate Professionals	<ISCED level 3A>	Heavy Truck & Lorry Drivers	Yes	One	30 minutes or less a day	Several times a week	Strongly disagree	15.42	0.8605	0.5091	524.08	567.74	550.58	24.0007	HUN: grammar school
9	00071	01780	10	Yes	Other Teaching Professionals Nec	<ISCED level 3A>	Higher Education Teaching Professionals [Incl. Univ. Professor]	Yes	Two	1 to 2 hours a day	A few times a year	Disagree	16	0.8605	-0.4352	490.19	499.5	503.95	16.5571	HUN: grammar school
10	00178	04458	10	Yes	Primary Education Teaching Professionals	<ISCED level 3A>	Primary Education Teaching Professionals	Yes	One	More than 2 hours a day	A few times a year	Strongly disagree	16	-0.108	0.5091	647.93	616.14	633.57	13.9604	HUN: grammar school
11	00065	01621	9	Yes	Primary Education Teaching Professionals	<ISCED level 3A>	[Small Enterprise] General Managers Wholesale & Retail Trade	Yes	Two	30 minutes or less a day	Several times a week	Disagree	15.83	0.8605	0.5091	499.54	535.25	494.63	21.2599	HUN: vocational secondary school
12	00130	03243	9	Yes	Do Not Know	<ISCED level 2>	Cabinet-Makers Etc Workers [Incl. Cartwright, Cooper]	Yes	One	30 minutes or less a day	A few times a year	Disagree	15.33	0.8605	-0.4352	478.9	498.31	497.71	17.6506	HUN: vocational secondary school
13	00010	00261	9	Yes	Finance & Sales Associate Professionals Nec	<ISCED level 3A>	Shop Salespersons & Demonstrators	Yes	Two	30 minutes or less a day	About once a month	Strongly disagree	15.67	0.8605	1.0878	520.57	477.69	509.83	27.1997	HUN: grammar school
14	00063	01562	9	Yes	Precision-Inst Makers & Repairers [Incl. Dental, Watch Maker]	<ISCED level 2>	Painters Etc Workers [Incl. Construction Painter, Paperhanger]	Yes	One	I don't read for enjoyment	A few times a year	Disagree	15.75	-0.108	0.0185	485.05	453	549.09	22.6412	HUN: vocational secondary school
15	00174	04359	11	Yes	Dentists	<ISCED level 3A>	Dentists	Yes	Two	30 minutes or less a day	About once a month	Strongly disagree	15.83	-0.108	2.0085	643.96	642.43	653.25	18.1435	HUN: grammar school
16	00120	03007	9	Yes	Production Dep. Managers Wholesale & Retail Trade	<ISCED level 3A>	Electrical Engineers	Yes	One	Between 30 and 60 minutes	Never or almost never	Disagree	15.25	0.8605	-0.821	509.59	564.64	550.86	27.7043	HUN: grammar school

The above file mainly consists of plain data.

More concepts here, ESS

I d e n t i f i e r (s) Survey questions Para data Weights

	Title of dataset	ESS round	Edition	Production date	Responden identificatio number	Country	Most people can be trusted or you can't be too careful	Boycotted certain products last 12 months	How happy are you	Discriminat of responden group: language	Citizen	Gende	Age of responden calculated	Children living at home or not at f4	Highest level of education	Worry about work problems when not working, how often	Usual gross pay in euros, before deductions for tax and insurance	Interviewer code: respondent lives with husband/wif	End of interview day of month	End of intervie month	Interview length in minutes, main questionnai	Day of month, supplemen questionna	Design weight	Population size weight (must be combined with dweight)
1	ESS5e02	5	2.0	28.03.2012	11380	FI	7	1	9	0	66	2	77	2	520	6	6666666	2	2	11	58	2	1.00	0.24
2	ESS5e02	5	2.0	28.03.2012	1701	FR	4	1	8	0	66	1	35	1	313	3	1690	1	26	11	61	26	1.01	3.05
3	ESS5e02	5	2.0	28.03.2012	12925	FI	7	2	8	0	66	2	95	2	113	6	6666666	2	5	10	87	5	1.00	0.24
4	ESS5e02	5	2.0	28.03.2012	4499	RU	8	2	9	0	66	2	75	2	520	6	6666666	2	18	2	82	18	0.57	4.64
5	ESS5e02	5	2.0	28.03.2012	132	BG	0	2	6	0	66	2	54	1	610	3	153	1	16	1	55	16	1.23	0.27
6	ESS5e02	5	2.0	28.03.2012	1238	NO	8	1	9	0	66	2	35	1	321	2	38462	1	20	10	64	20	1.19	0.25
7	ESS5e02	5	2.0	28.03.2012	11666	FI	8	2	7	0	66	2	46	1	520	4	3300	2	13	10	52	13	1.00	0.24
8	ESS5e02	5	2.0	28.03.2012	10333	HU	6	2	4	0	66	2	17	2	213	6	6666666	2	6	11	60	6	0.99	0.55
9	ESS5e02	5	2.0	28.03.2012	21771300	DE	0	1	2	0	66	1	59	2	322	6	6666666	2	28	9	116	28	0.53	2.34
10	ESS5e02	5	2.0	28.03.2012	966	PT	4	2	7	0	66	2	72	2	520	6	6666666	1	10	2	66	.	0.31	0.42
11	ESS5e02	5	2.0	28.03.2012	10367	CY	5	2	8	0	66	2	31	1	710	4	7777777	1	4	2	75	4	0.70	0.06
12	ESS5e02	5	2.0	28.03.2012	2147	RU	4	2	7	0	66	1	18	2	313	6	6666666	2	17	1	79	17	1.89	4.64
13	ESS5e02	5	2.0	28.03.2012	625	ES	4	2	6	0	66	2	63	2	620	6	6666666	1	4	5	66	4	1.01	2.08
14	ESS5e02	5	2.0	28.03.2012	2450	ES	4	2	8	0	66	2	34	2	720	6	6666666	1	11	6	69	11	1.02	2.08
15	ESS5e02	5	2.0	28.03.2012	503166	NL	6	2	9	0	66	1	37	1	520	3	3100	1	8	11	66	8	1.01	0.75
16	ESS5e02	5	2.0	28.03.2012	2394	BG	0	2	9	0	66	2	40	1	213	6	6666666	1	22	1	65	22	1.71	0.27
17	ESS5e02	5	2.0	28.03.2012	2493	HR	5	2	7	0	66	1	21	2	313	6	6666666	2	23	9	60	23	1.11	0.23
18	ESS5e02	5	2.0	28.03.2012	17207	GR	5	2	2	0	66	2	29	1	520	6	6666666	1	22	6	65	22	0.88	0.36
19	ESS5e02	5	2.0	28.03.2012	2290	HR	6	2	4	0	66	2	42	1	323	4	4000	1	19	9	60	19	0.77	0.23
20	ESS5e02	5	2.0	28.03.2012	202925	DK	8	1	9	0	66	2	32	1	520	2	8888888	1	29	10	43	29	1.00	0.29
21	ESS5e02	5	2.0	28.03.2012	683	ES	5	2	8	0	66	2	38	1	720	2	7777777	1	11	5	66	11	1.17	2.08
22	ESS5e02	5	2.0	28.03.2012	501505	NL	6	2	9	0	66	2	34	1	620	4	3000	1	31	3	71	31	1.01	0.75
23	ESS5e02	5	2.0	28.03.2012	2759	IL	8	1	7	0	66	2	49	1	720	3	12500	2	27	1	73	27	1.32	0.24
24	ESS5e02	5	2.0	28.03.2012	30440713	GB	5	2	7	0	66	2	70	2	113	6	6666666	2	27	9	56	27	0.53	2.11
25	ESS5e02	5	2.0	28.03.2012	2442	HR	7	1	7	0	66	2	55	2	520	6	6666666	1	29	10	45	29	1.11	0.23
26	ESS5e02	5	2.0	28.03.2012	102	CZ	3	2	7	0	66	2	36	2	321	2	7777777	1	31	1	125	31	0.99	0.38
27	ESS5e02	5	2.0	28.03.2012	106600	DK	9	2	9	0	66	2	60	2	213	6	6666666	1	26	10	42	26	1.00	0.29
28	ESS5e02	5	2.0	28.03.2012	1810	GR	6	2	7	0	66	2	50	1	213	1	1100	1	11	6	60	11	1.77	0.36
29	ESS5e02	5	2.0	28.03.2012	24528871	GB	3	2	7	0	66	1	53	2	113	6	6666666	1	5	10	67	5	1.06	2.11
30	ESS5e02	5	2.0	28.03.2012	22638	IE	2	2	3	0	66	2	58	2	213	2	14000	2	.	.	.	99	0.47	0.14
31	ESS5e02	5	2.0	28.03.2012	2304	NO	7	2	5	0	66	1	34	1	720	2	32051	1	11	10	63	11	0.77	0.25
32	ESS5e02	5	2.0	28.03.2012	10202	GR	3	2	4	0	66	2	79	2	113	6	6666666	2	28	6	60	28	0.44	0.36
33	ESS5e02	5	2.0	28.03.2012	2585	PL	5	2	8	0	66	2	80	2	229	6	6666666	1	9	10	80	9	0.86	1.85
34	ESS5e02	5	2.0	28.03.2012	21953000	DE	10	2	10	0	66	1	17	2	213	6	6666666	2	22	9	71	22	1.30	2.34
35	ESS5e02	5	2.0	28.03.2012	500040	NL	4	2	10	0	66	2	71	2	520	6	6666666	1	17	3	55	17	1.01	0.75
36	ESS5e02	5	2.0	28.03.2012	597	UA	5	2	3	0	66	1	56	2	720	1	2000	2	26	6	63	26	0.29	2.04
37	ESS5e02	5	2.0	28.03.2012	503	IL	7	2	8	0	66	1	60	2	620	3	9999999	1	8	5	.	8	1.23	0.24
38	ESS5e02	5	2.0	28.03.2012	2406	EE	8	1	8	0	66	2	32	1	423	6	6666666	1	.	.	.	25	1.00	0.06
39	ESS5e02	5	2.0	28.03.2012	502414	NL	3	2	2	0	66	1	50	2	113	6	6666666	2	18	1	55	18	0.51	0.75
40	ESS5e02	5	2.0	28.03.2012	1459	IL	5	2	5	0	66	1	31	2	313	2	6000	1	1	2	60	1	1.10	0.24
41	ESS5e02	5	2.0	28.03.2012	14205	GR	3	2	6	0	AL	2	31	1	113	6	6666666	1	4	6	55	4	0.88	0.36
42	ESS5e02	5	2.0	28.03.2012	378	ES	8	2	8	0	66	1	51	1	222	4	8888888	1	25	7	43	25	1.04	2.08
43	ESS5e02	5	2.0	28.03.2012	16318	IE	8	2	8	0	66	1	65	2	313	6	6666666	1	.	.	.	26	0.94	0.14
44	ESS5e02	5	2.0	28.03.2012	22124114	GB	7	2	8	0	66	2	24	2	720	5	36270	2	13	9	40	13	0.53	2.11
45	ESS5e02	5	2.0	28.03.2012	200001632	SE	8	2	10	0	66	1	19	2	323	1	1720	2	3	11	44	3	1.00	0.52
46	ESS5e02	5	2.0	28.03.2012	9044	RU	2	2	5	0	66	2	34	1	313	1	500	2	4	3	45	4	0.35	4.64
47	ESS5e02	5	2.0	28.03.2012	100000696	SE	4	2	8	0	66	2	45	1	313	6	6666666	1	3	11	66	3	1.00	0.52
48	ESS5e02	5	2.0	28.03.2012	1260	CZ	6	2	8	0	66	1	25	2	321	6	6666666	2	24	2	120	24	1.50	0.38
49	ESS5e02	5	2.0	28.03.2012	2588	IL	5	2	5	0	66	1	87	2	620	6	6666666	2	13	4	60	13	0.41	0.24
50	ESS5e02	5	2.0	28.03.2012	17935	IE	3	2	88	0	66	1	19	2	313	6	6666666	2	.	.	.	1	1.41	0.14
51	ESS5e02	5	2.0	28.03.2012	1907	NO	2	2	8	0	66	2	63	2	313	3	48718	1	26	10	74	26	0.94	0.25
52	ESS5e02	5	2.0	28.03.2012	1777	RU	4	2	4	0	66	1	22	2	313	6	6666666	2	19	3	45	19	1.28	4.64

Cleaned survey data

Next we go forward to operations that starts for designing data collection strategies and then getting raw data. The raw data need to be cleaned, as well.

Some more is needed:
-You have to document everything somewhere, and most important things into the electronic file, that is,
- You have to label your variables
- You have to label the classifications of your variables
- You have to add para data into the file; this is mainly derived from the fieldwork including e.g.
    - Interviewing time and place, length, interviewer code. mode
    - Reasons for missingness
    - Comments on data quality
- You have to save your file in a good format

And if you are releasing your data set outsiders
- You have to make the data confidential.