(Version 1.1)

Users' Manual for Handling Resampled Micro Data of Vietnamese Household Living Standard Survey (VHLSS)

VHLSS – Overall and Survey Process

2015 The Institute of Statistical Mathematics (ISM) and Statistical Information Institute for Consulting and Analysis (SINFONICA)

History of revision of the manual

□ Version 1.1 in February 2015

- Revised based on the discussion during the Sixth International Workshop on Analysis of Micro Data of Official Statistics in December 2014

- Added chapter 6.8 Education

□ First draft version 1.0 in September 2014

CONTENTS

VHLSS – Overall and survey process	
1. About this Manual	Page 3
2. Outline of VHLSS	
2.1 Objective of the survey	4
2.2 Topics covered by the survey	4
2.3 Coverage of the survey	5
2.4 Sample design	6
2.5 Data collection and editing	10
2.6 Data processing	11
3. Data and metadata provided	
3.1 VHLSS 2002/2004/2006	12
3.2 Other source of metadata	33
4. Map of provinces and regions	38
5. Panel data (Trial)	40
6 Population and household	
6.1 Ethnicity	66
6.2 Pelicion	68
6.3 Age bearing	69
6.4 Literacy	71
6.5 Polygamy	74
6.6 Relationship to the household head	78
6.7 Family type	
6.7.1 Family type based on the number of persons by generation	84
6.7.2 Family type based on the generation and number of married couples	94
6.8 Education	100
7. Agricultural land	103

2

1. About this Manual

1. This manual was prepared for users to use the next <u>80% resampled micro data</u> sets of Vietnamese Household Living Standard Survey (VHLSS). It described mainly overall and survey process of VHLSS 2002, 2004 and 2006.

Survey year	Data files	Contents
VHLSS 2002	31 files	Household questionnaire for Income and Expenditure Survey
VHLSS 2004	47 files	Household questionnaire for Income and Expenditure Survey
VHLSS 2006	49 files	Household questionnaire for Income and Expenditure Survey

2. The questionnaires, file layout and data dictionary of each VHLSS are described in the manual for each year's VHLSS, respectively.

3. The original micro data sets composed of all the samples were provided by NSO, Viet Nam based on the Charter for Experimental Laboratory for Research Purpose Statistical Use of Micro Data, and resampled at the rate of 80% by Sinfonica.

4. This manual was first compiled in September 2014 by;Hiroshige FurutaVisiting Senior Research Fellow, Sinfonica

Acknowledgements

Special thanks to Ms. Van Nghiem, GSO, Vietnam, who assisted my work of compiling the manuals by properly answering to my queries via email.

2. Outline of VHLSS

This chapter mainly describes on VHLSS 2002, 2004 and 2006.

2.1 Objective of the survey

In order to evaluate the living standards for policy and socio-economic plan making, the General Statistics Office (GSO) has conducted many household living standards surveys. From 2002 to 2010 particularly, the household living standards survey (VHLSS) is conducted every 2 years by the GSO in the years end with even numbers in order to monitor systematically living standards of Vietnam population's groups; monitor and assess the implementation of the Comprehensive Poverty Reduction and Growth Strategy; making contribution to evaluating results of realization of the Millennium Development Goals (MDGs) and Vietnam Development Goals (VDGs).

The GSO publishes the official results of the VHLSS. Data of other statistical areas compiled from the VHLSS are not for replacing data of these areas which were published before rather than for clarifying and further analyzing factors affecting to living standards.

VHLSS collects information to be used as basis for assessment of living standard, including poverty and the gap between the rich and the poor serving for policy making, planning and national targeted programs of the party and the State in order to continuously improve the living standard of population across the country, in all regions and localities.

In addition to that, information is collected to serve for research, analysis of some topics on health, education, employment and to calculate weight to compile consumer price index and national account.

2.2 Topics covered by the survey

VHLSS includes main content reflecting living standard of households in the entire country, and main socio-economic conditions of communes in the rural areas which affect living standard of population in their particular area.

VHLSS consists of two surveys; for household and for commune.

2.2.1 For households

- Some demographic characteristics of household's members including age, sex, ethnicity, marital status.

- Household's income includes: income level; income by different sources (salary, wage; self-agricultural, forestry, fishery production; self-household's business production/service; others);

income by economic sector and industry.

- Household's expenditure: expenditure level, expenditure by purpose and item (expenditure on food, clothes, accommodation, travel, education, health, culture, etc... and others.

- Education level, professional level of each household's member.

- Illness and use of type of health clinic.

- Employment status, working hours.

- Assets, housing and facilities, such as appliances, electricity, water, sanitation condition.

- Participation in hunger elimination and poverty reduction, credit status.

If budget is available, extended content on education and health will be surveyed.

2.2.2 For communes which have surveyed households

- Some information on demography, ethnicity.

- Basic socio-economic infrastructure including: situation of electricity, roads, schools, medical stations, markets, post offices, water sources.

- Economic situation, including: Agriculture production (land, production increase/decrease tendency of some main crops and causes, assistances for production development such as irrigation, agricultural encouragement); non-agricultural job opportunities.

- Some main information on social order and safety.

Note: Micro data of the survey for commune was not provided.

2.3 Coverage of the survey

2.3.1 VHLSS covers the whole country. Scope of the survey includes all selected enumeration areas and communes in 64 provinces and cities (in VHLSS 2004 and 2006) under central management (hereafter called province/city).

2.3.2 Target population

The target population comprises the civilian, non-institutional population.

Only persons considered as permanent residents are eligible for inclusion in the survey.

The following persons are included;

-persons on vacation

-temporarily in a hospital

-students living away from home during the school year

The following households are ineligible;

-housing unit containing only students living away from home during the school year
-military compounds
-hospitals
-prisons

2.4 Sample design: three-stage stratified cluster design

Sample of the VHLSS is selected in the way to represent the entire country (in which: urban/rural areas), 8 regions (in which: urban/rural areas), and provinces/cities.

Master sample

Two-stage area sample from enumeration areas of the 1999 Population and Housing Census.

Strata: province and urban/rural

PSU (primary sampling unit): communes

Out of 10,476 communes as of 1999 Population Census, 3,063 communes/words were selected.

Communes contain on average about 1,600 households while EAs were set up during the censuses to contain about 100 households.

SSU (Secondary sampling unit): EAs

Three EAs were selected per commune. Only one of EA was used for each year of VHLSS survey. This is technically a three-stage design counting the selection of households, but it is operationally equivalent to a two-stage design.

Sample allocation

The sample was allocated over strata proportional to the square root of the number of households.

Both communes and EAs were selected with probability proportionate to size (PPS), the size being the number of households according to 1999 Population Census.

The master sample is designed specifically for VHLSSs in the period 2002-2010. It contains 3,063 communes/wards, and 3 EAs are selected in each commune/ward from EAs of the Population and Housing Census 1999.

6

Figure 1 Structure of VHLSS 2006

Type of survey	Target and size (2006)	Questionnaire/Questions	Objective
Commune survey	3,063 commune/wards	Basic socio-economic	
	which have surveyed	information	
	households		
Income survey	36,756 households	(Short) Questionnaire 1A;	To assess living
		Questions on income	standards at
			provincial/city level
Income and	9,189 households	(Long) Questionnaire 1B;	To assess living
expenditure		Questions on income and	standards at national
survey		expenditure	and regional level
In total	45,945 households		

Sample size

Sample size of the VHLSS 2006 includes 45,945 households selected from 3,063 areas of the master sample frame, and is divided into 2 types: *Sample for income survey* includes 36,756 households to collect information as mentioned above, excluding household's expenditure, to assess living standard at national level, regional and provincial/city level; *Sample for income-expenditure survey* includes 9,189 to collect sufficient information for further assessment and analysis of living standard <u>at national and regional level</u>.

Sample for income survey and sample for income-expenditure survey was divided for two rounds are as follows:

Time for	Survey on income	Survey on	Total
Data collection	& expenditure	income	
Total	9.189	36.756	45.945
Of which:	(households)		
May 2006	4.593	18.372	22.965
September 2006	4.596	18.384	22.980

Sample selection for each year of VHLSS: 50% rotation of households

The Socio-Environmental Statistics Department coordinates with provinces/cities to select sample as follows:

Step 1: Select EAs.

EAs of the VHLSS 2006 will be selected rotationally, specifically: <u>re-select 50%</u> <u>enumeration areas of the VHLSS 2004</u> (in which half of the areas were surveyed in the 2002 and 2004 VHLSSs and another half of the areas were only surveyed in the 2004 VHLSS) and the other 50% areas will be newly selected from the master sample, which were not selected in the 2002 and 2004 VHLSSs.





Panel data

The sample design enables to make use of panel data. The questionnaire of VHLSS 2004 and 2006 included questions about the identification code of the previous VHLSS, which were recorded by the enumeration team leader.

Remarks:

Identifiers of VHLSS 2002 differ from the later VHLSS, as follows.

VHLSS 2004/2	006	VHLSS 2002	,
tinh	3	tinh	
huyen	2		xa
xa	2		
diaban	3		hase
hoso	2		noso

Updating of sample frame of households

The Socio- Environmental Statistics Department is responsible for selecting areas and sending the list of selected areas to Provincial Statistics Offices for reviewing and updating attached with the map and list of areas of the 1999 Population and Housing Census of the new areas.

Provincial Statistics Offices review and propose to change some areas for more suitability with geographical, socio-economic characteristics of provinces with less than the change of 5% of total number of provinces' areas with an agreement of the GSO (the Socio-Environmental Statistics Department) during the year prior to implementing the survey.

Step 2: Select households.

Provincial Statistics Offices select households, specifically:

- <u>For areas which are re-selected from the 2004 VHLSS</u>, select all 15 households in which 12 households were already surveyed with income (income households) in 2004 to be surveyed with income for the 2006 VHLSS and <u>3 households were already surveyed with income-expenditure in 2004 to be surveyed with income-expenditure for the 2006 VHLSS</u>. In case of households which were surveyed in 2002 or 2004 moved from the area, <u>find alternate households to be assured of</u> 12 households for income and <u>3 households for income-expenditure in each enumeration area</u>.

- <u>For new areas</u>, select 20 households from the updated household's list. From these 20 households, select 15 households (12 official households, 3 spare households) for income survey, the 5 remaining households (3 official and 2 spare households) for income-expenditure survey.

Household selection follows the methods in the 2006 VHLSS manual. Provincial Statistics Office will equally divide EAs by urban/rural areas and geographical region for the two rounds of the surveys conducted in May and September.

Communes which have selected EAs for the household interview will simultaneously interview commune questionnaire.

The list of selected households will be kept in two places: PSOs and Socio-environmental Statistics Department for the implementation, monitoring, and supervision.

Remarks:

Sample design was changed based on the results of analysis of VHLSS 2002. In VHLSS 2002, the cluster size (the number of sample households within EA) was 25 households. The cluster sizes for the expenditure module in VHLSS 2004 and 2006 were smaller, 3 households per EA as compared to 5 or 20 households per EA in VHLSS 2002.

Definition of urban/rural

The definition of urban/rural is not clear.

There are three types of local administrative unit; ward, district town and commune. They are classified into urban and rural.

Urban: ward and district town Rural: commune

Note: The area of urban/rural might vary year by year.

2.5 Data collection method

Field work

The field work was done about one month for each round of the survey.

In VHLSS 2002, it was implemented in 4 quarters (starting at first month of each quarter). Each long questionnaire was completed within one and a half day, and short questionnaire was completed within one day.

In VHLSS 2004, the survey was implemented in May and September. The long questionnaire (including two special modules) was often completed within 2.5 days and the short is completed in one and a half.

Questionnaire

This survey uses two types of questionnaire: questionnaire for commune and questionnaire for household, as in Figure 1.

The questionnaire for household includes: income-expenditure questionnaire (applied for the sample of the income-expenditure survey) which contains all information of the survey; and income questionnaire (applied for sample of the income survey) which contains all information of the survey excluding information on household's expenditure. Questionnaires are designed in details so that it will facilitate the recording of interviewers, help avoid the missing of items and increase the consistency among interviewers. Thus the survey data will be improved.

Direct interview is applied in this survey. Interviewers visit households, meet with heads of households and related households' members to interview and fill information in questionnaires for households.

2.6 Data processing

1) The data is first checked by team leader before sending to PSO.

2) The PSO checks the data again and send it to data entry staffs.

3) The data entry staffs uses CSPro for entering data twice and data cleaning. The program gives a warning for inconsistent or out of range for each suspicious numbers in each question.

4) After the data entry is completed at PSO, it is sent to Center of Information and Statistics as well as to Social and Environment Statistics Department. The second data cleaning process is done at the Social and Environment Statistics Department. Total time for data cleaning often takes from 8 to 12 months.

[VHLSS 2006]

After finishing data entering and cleaning up, Provincial Statistics Offices have to preliminarily compile the survey data. Leaders from Provincial Statistics Offices examine and send the preliminarily compiled data and data testing report to the Socio- Environmental Statistics Department for assessment. The compilation and assessment of the preliminary results should be finished in 1 month, of them 20 days are for preliminary results and 10 days are for assessment of the preliminary results.

After the preliminary results have been assessed, Provincial Statistics Offices proceed to officially compilation, as well as send the entered raw data to Hanoi Statistical Informatics Center.

The Statistical Computing Centre of Hanoi chairs the cooperation with Socio-Environmental Statistics Department to compile the nationwide data.

The General Statistics Office publishes the preliminary results of the VHLSS 2006 in June 2007 and official results in December 2007.

3 Data and metadata provided by NSO

3.1 Viet Nam Household Living Standard Survey 2002/2004/2006

The data and metadata were provided to Sinfonica in February 2012 by Ms. Ngheim Thi Van, GSO official cum TIU student at that time, with the permission of NSO.

In the above data set, the weight data had not been included. The weight data files were submitted to Sinfonica by Ms. Van, GSO official in July 2014 upon request.

Data and metadata provided

Organization of data files

Summary sheet.xlsx	Description of list of data f	iles including the next items;
	Year, STATA filenames,	Section in the questionnaire, Main contents,
	Number of variables	
	Sheet '2002'	Number of data files is 31.
		Number of sections is 9.
	Sheet '2004'	Number of data files is 47.
		Number of sections is 9.
	Sheet '2006'	Number of data files is 49.
		Number of sections is 8.

Example of Sheet '2006'

	A	В	С	D	E	F
1				List of data files		
			STATA			Number of
2	Year	No.	filenames	Section	Main contents	variables
3	2006	1	muc1a.dta	Section 1: List of household members	List of household members	17
4	2006	2	muc1b.dta	Section 1: List of household members	List of household members	16
5	2006	3	muc2a.dta	Section 2: Education, training and vocational training	General information	34
6	2006	4	muc2b.dta	Section 2: Education, training and vocational training	Detail information on general and tertiary education	21

			VHLSS 2002 - List of c	lata files	
Year	No.	STATA	Section	Main contents	Number of
		filenames			variables
2002	1	muc1	Section 1: List of household members	List of household members	13
2002	2	muc2	Section 2: Education	Education level and expense of household members	25
2002	3	muc2ho	Section 2: Education	Total expense for education of household	9
2002	4	muc3	Section 3: Employment	Employment	27
2002	5	muc4	Section 4: Health	Health situation and expense of household members	11
2002	6	muc4ho	Section 4: Health	Total expense for health of household	14
2002	7	muc5a	Section 5: Income and other inflows of money	income from salary, wage of household members	18
2002	8	muc5aho	Section 5: Income and other inflows of money	Total income from salary, wage of household	10
2002	9	muc5b1	Section 5: Income and other inflows of money	Agricultural, forestry and fishery activities	68
2002	10	muc5b2ho	Section 5: Income and other inflows of money	Agriculture production, Income from Annual and perennial industrial	26
				crops, Income from Fruit crops, Income from Crop by-production,	
				Agricultural production expenditure	
2002	11	muc5b4	Section 5: Income and other inflows of money	Farm services, Income from Farm services, Expenses for Farm	88
				services of household	
2002	12	muc5b5	Section 5: Income and other inflows of money	Sylviculture, Income from sylvicuture, Sylvicultural expenditure	62
2002	13	muc5b21_23	Section 5: Income and other inflows of money	Income from rice and other food crops, Income from Annual and	13
				perennial industrial crops, Income from Fruit crops of household	
2002	14	muc5b25	Section 5: Income and other inflows of money	Agricultural production expenditure	11

2002	15	muc5b31	Section 5: Income and other inflows of money	Income from livestock	108
2002	16	muc5b32	Section 5: Income and other inflows of money	Expenditure on livestock	72
2002	17	muc5b61	Section 5: Income and other inflows of money	income from aquaculture	73
2002	18	muc5b62	Section 5: Income and other inflows of money	Expenditure on planting and growing aquacultural products	49
2002	19	muc5c1	Section 5: Income and other inflows of money	Income from non-farm, non-forestry and non-aquaculture businesses	43
2002	20	muc5c2	Section 5: Income and other inflows of money	EXPENDITURE FOR non-farm, non-forestry and non-aquaculture	55
				businesses	
2002	21	muc5d	Section 5: Income and other inflows of money	Other sources of income	27
2002	22	muc6a1	Secton 6: Expenditure	expenditure on food and drinks during holidays	10
2002	23	muc6a2	Secton 6: Expenditure	Daily Expenditure on food and drink	15
2002	24	muc6b1	Secton 6: Expenditure	Expenditure on daily consumption	10
2002	25	muc6b2	Secton 6: Expenditure	annual expenditure	104
2002	26	muc6b34	Secton 6: Expenditure	other expenses included and not included in expenditure	30
2002	27	muc7	Section 7: Fixed assets and durable things	fixed assets and durable things	13
2002	28	muc7ho	Section 7: Fixed assets and durable things	fixed assets and durable things	9
2002	29	muc8	Section 8: Housing	housing	120
2002	30	muc9	Section 9: Participation in poverty alleviation	Participation in poverty alleviation programs	49
			programs		
2002	31	tongcong	Summary	summary	58

			VHLSS 2004 - List of	f data files	
Year	No	STATA	Section	Main contents	Number of
		filenames			variables
2004	1	M1_2_3A	Section 1,2,3: List of household members,	list of household members, Education, training and vocational	49
			Education and Health	training, health and health care	
2004	2	M1B	Section 1: List of household members	list of household members	14
2004	3	M3B	Section 3: Health and health insurance	Health and health care and health insurance	17
2004	4	M4A	Section 4: Income	Employment	45
2004	5	M4B11	Section 4: Income	Rice	20
2004	6	M4B12_14	Section 4: Income	other starchy, vegetable, annual and perennial industrial crops, Fruit	14
				Crops	
2004	7	M4B15	Section 4: Income	income from crop by-products	11
2004	8	M4B16	Section 4: Income	Crop Planting Expenditure	12
2004	9	M4B21	Section 4: Income	Income from Livestock breeding	15
2004	10	M4B22	Section 4: Income	Livestock breeding expenditure	19
2004	11	M4B31	Section 4: Income	Income from Agricultural Services	10
2004	12	M4B32	Section 4: Income	Expenses on Agricultural Services	18
2004	13	M4B41	Section 4: Income	Income from forestry and hunting, trapping and domesticating forest	9
				animals and birds	
2004	14	M4B42	Section 4: Income	expenditure for forestry and hunting, trapping and domesticating	21
				forest animals and birds	
2004	15	M4B51	Section 4: Income	income from aquaculture	15

2004	16	M4B52	Section 4: Income	expenditure for aquaculture	20
2004	17	M4C1	Section 4: Income	Non-agriculture activity	25
2004	18	M4C2	Section 4: Income	expenditure for from non-agriculture, non-forestry, non-aquaculture	19
				business and production trades; the process of agricultural,	
				forestry, and aquacultural products	
2004	19	M4D	Section 4: Income	Other income	20
2004	20	M5A1	Section 5: Expenditure	Expenditure on foods and drinks during holidays	11
2004	21	M5A2	Section 5: Expenditure	Daily expenditure on foods and drinks	16
2004	22	M5B1	Section 5: Expenditure	Expenditure on daily consumption	11
2004	23	M5B2	Section 5: Expenditure	Annual consumption expenditure	9
2004	24	M5B3_4	Section 5: Expenditure	Other spending that is considered and is not considered as	22
				household expenditure,	
2004	25	M6	Secton 6: Fixed assets and consumer durables	Fixed assets and durable appliances	68
2004	26	M6A	Secton 6: Fixed assets and consumer durables	Fixed assets and durable appliances	13
2004	27	M6B	Secton 6: Fixed assets and consumer durables	Fixed assets and durable appliances	12
2004	28	M7	Section 7: Housing, water and sanitation	Accommodation	52
2004	29	M8	Section 8: Participation in the poverty	Participation in the poverty alleviation and hunger eradication	42
			alleviation and hunger eradication programme	program	
			and credit		
2004	30	M8_XDGN	Section 8: Participation in the poverty	Participation in the poverty alleviation and hunger eradication	19
			alleviation and hunger eradication programme	program	
			and credit		

2004	31	M10A_E3	Section 10: Business other than agriculture,	Information on time, location, and labor, Business history,	89
			forestry and aquaculture (expanded)	involvement in business associations and clubs, Contact with	
				relevant agencies at different levels and industries, other	
				characteristics	
2004	32	M10F	Section 10: Business other than agriculture,	Occupations, industries which existed during the past 10 years but	15
			forestry and aquaculture (expanded)	ceded operations over the12 months	
2004	33	M91	Section 9: Agriculture, forestry and aquaculture	FARMING, FORESTRY LAND AND WATER SURFACE FOR	23
			(expanded)	AQUACULTURE, residential LAND AND GARDEN, POND NEXT	
				TO HOUSING LAND	
2004	34	M92	Section 9: Agriculture, forestry and aquaculture	land that is rented/borrowed/temporarily exchanged	14
			(expanded)		
2004	35	M93	Section 9: Agriculture, forestry and aquaculture	Land that is temporaily rented out/ lent out	14
			(expanded)		
2004	36	M941	Section 9: Agriculture, forestry and aquaculture	rice	11
			(expanded)		
2004	37	M942_44	Section 9: Agriculture, forestry and aquaculture	staple food, food crops and other annual crops, annual and	10
			(expanded)	perennial industrial crops, fruit crops	
2004	38	M946	Section 9: Agriculture, forestry and aquaculture	access to extension services	11
			(expanded)		
2004	39	M948	Section 9: Agriculture, forestry and aquaculture	Bought, bid, inherited land or use right-transferred land over the last	19
			(expanded)	10 years	
2004	40	M949	Section 9: Agriculture, forestry and aquaculture	Sold, tender-expired, acquired land or inheritance right-transfered	20
			(expanded)	land in the last 10 years	

2004	41	M9451	Section 9: Agriculture, forestry and aquaculture	Cropping structure	10
			(expanded)		
2004	42	M9452	Section 9: Agriculture, forestry and aquaculture	Cropping structure	15
			(expanded)		
2004	43	M9453	Section 9: Agriculture, forestry and aquaculture	Cropping structure	13
			(expanded)		
2004	44	M9471	Section 9: Agriculture, forestry and aquaculture	conversions in agricultural, forestry and aquacultural land and over	17
			(expanded)	the past 10 years	
2004	45	M9472	Section 9: Agriculture, forestry and aquaculture	conversions in agricultural, forestry and aquacultural land and over	15
			(expanded)	the past 10 years	
2004	46	ho1	Summary	Summary	128
2004	47	ho2	Summary	Summary	39

VHLSS 2006 - List of data files						
Year	No.	STATA	Section	Main contents	Number of	
		filenames			variables	
2006	1	muc1a.dta	Section 1: List of household members	List of household members	17	
2006	2	muc1b.dta	Section 1: List of household members	List of household members	16	
2006	3	muc2a.dta	Section 2: Education, training and vocational training	General information	34	
2006	4	muc2b.dta	Section 2: Education, training and vocational training	Detail information on general and tertiary education	21	
2006	5	muc2c.dta	Section 2: Education, training and vocational training	Extra classes	31	
2006	6	muc2d.dta	Section 2: Education, training and vocational training	Repeated the school year	19	
2006	7	muc2e.dta	Section 2: Education, training and vocational training	Vocational training	16	
2006	8	muc3a1.dta	Section 3: Health and health insurance	Health and health care	12	
2006	9	muc3a2.dta	Section 3: Health and health insurance	Health and health care	18	
2006	10	muc3b.dta	Section 3: Health and health insurance	Disability	93	
2006	11	muc3c.dta	Section 3: Health and health insurance	Health	17	
2006	12	muc3d.dta	Section 3: Health and health insurance	Fertility	15	
2006	13	muc3e.dta	Section 3: Health and health insurance	Behavoriours that have impacts on health	11	
2006	14	muc3f.dta	Section 3: Health and health insurance	Health insurance	28	
2006	15	muc3g.dta	Section 3: Health and health insurance	Out-patient diagnosis treatment	28	
2006	16	muc3h.dta	Section 3: Health and health insurance	In-patient diagnosis treatment	31	
2006	17	muc3i.dta	Section 3: Health and health insurance	Self-treatment	16	
2006	18	muc4a.dta	Section 4: Income	Employment	52	
2006	19	muc4b0.dta	Section 4: Income	Land for agriculture, forestry and water surface for	16	

				aquaculture	
2006	20	muc4b11.dta	Section 4: Income	Rice	19
2006	21	muc4b12.dta	a Section 4: Income Other starchy, vegetable		12
2006	22	muc4b13.dta	Section 4: Income	Annual and perennial industrial crops	13
2006	23	muc4b14.dta	Section 4: Income	Fruit crops	13
2006	24	muc4b15.dta	Section 4: Income	Income from crops by products	10
2006	25	muc4b16.dta	Section 4: Income	Crop planning expenditure	11
2006	26	muc4b161.dta	Section 4: Income	Table of quantity of chemical fertilizers used for types of	11
				trees	
2006	27	muc4b21.dta	Section 4: Income	Income from livestock breeding	14
2006	28	muc4b22.dta	Section 4: Income	Livestock breeding expenditure	18
2006	29	muc4b31.dta	Section 4: Income	Income from agricultural services	9
2006	30	muc4b32.dta	Section 4: Income	Expenses on agricultural services	17
2006	31	muc4b41.dta	Section 4: Income	Income from forestry and hunting, trapping and	13
				domesticating forest animals and birds	
2006	32	muc4b42.dta	Section 4: Income	Expenditure from forestry and hunting, trapping and	20
				domesticating forest animals and birds	
2006	33	muc4b51.dta	Section 4: Income	Income from aquaculture	14
2006	34	muc4b52.dta	Section 4: Income	Expenditure from aquaculture	19
2006	35	muc4c.dta	Section 4: Income	Non-agriculture, non-forestry, non-aquaculture business	33
				and production trades	
2006	36	muc4c2.dta	Section 4: Income	Expenditure on -nNon-agriculture, non-forestry,	9
				non-aquaculture business and production trades	

2006	37	muc4d.dta	Section 4: Income	Other income	22
2006	38	muc5a1.dta	Section 5: Expenditure	Expenditure on foods and drinks during holidays	10
2006	39	muc5a2.dta	Section 5: Expenditure	Daily expenditure on food and drinks	15
2006	40	muc5b1.dta	Section 5: Expenditure	Daily non-foods expenditure and other expenditures	10
2006	41	muc5b2.dta	Section 5: Expenditure	Annual consumption expenditure	8
2006	42	muc5b3_4.dta	Section 5: Expenditure	Other spending that is considered and not considered as	23
				household expenditure	
2006	43	muc6.dta	Section 6: Fixed assets and durable appliances	Fixed assets and durable appliances	67
2006	44	muc6a.dta	Section 6: Fixed assets and durable appliances	Fixed assets and durable appliances	12
2006	45	muc6b.dta	Section 6: Fixed assets and durable appliances	Fixed assets and durable appliances	11
2006	46	muc7.dta	Section 7: Accommodation	Accommodation	51
2006	47	muc8.dta	Section 8: Participation in the poverty alleviation and	Participation in the poverty alleviation and hunger	37
			hunger eradication programme	eradication programme	
2006	48	muc8_vayvon.dta	Section 8: Participation in the poverty alleviation and	Participation in the poverty alleviation and hunger	20
			hunger eradication programme	eradication programme	
2006	49	ttchung.dta	Summary	Summary	141

[VHLSS 2002]

Outline of the survey;

General Instruction of	Description of survey objectives, survey contents, survey methodology,
VHLSS2002.doc	training and field survey, main contents of VHLSS2002

Data in STATA and SAS format

File names in STATA format;								
[1] "muc1.dta"	"muc2.dta"	"muc2ho.dta"	"muc3.dta"					
[5] "muc4.dta"	"muc4ho.dta"	"muc5a.dta"	"muc5aho.dta"					
[9] "muc5b1.dta"	"muc5b21_23.dta	a" "muc5b25.dta"	"muc5b2ho.dta"					
[13] "muc5b31.dta"	"muc5b32.dta"	"muc5b4.dta"	"muc5b5.dta"					
[17] "muc5b61.dta"	"muc5b62.dta"	"muc5c1.dta"	"muc5c2.dta"					
[21] "muc5d.dta"	"muc6a1.dta"	"muc6a2.dta"	"muc6b1.dta"					
[25] "muc6b2.dta"	"muc6b34.dta"	"muc7.dta"	"muc7ho.dta"					
[29] "muc8.dta"	"muc9.dta"	"tongcong.dta"						
Number of records and	d variables in each d	lata file;						
muc1.dta : 132384 1	3							
muc2.dta : 132374 2	25							
muc2ho.dta : 29532	9							
muc3.dta : 109715 2	27							
muc4.dta : 29522 11								
muc4ho.dta : 29532	14							
muc5a.dta : 27220 1	muc5a.dta : 27220 18							
muc5aho.dta : 29532	2 10							
muc5b1.dta : 29530	68							
muc5b21_23.dta : 10	09865 13							
muc5b25.dta : 1535	65 11							
muc5b2ho.dta : 2953	32 26							
muc5b31.dta : 2953	0 108							
muc5b32.dta : 2953	0 72							
muc5b4.dta : 29530	88							
muc5b5.dta : 29530	62							
muc5b61.dta : 29530 73								
muc5b62.dta : 2953	muc5b62.dta : 29530 49							
muc5c1.dta : 29530	43							
muc5c2.dta : 29530	55							

muc5d.dta : 29530 27
muc6a1.dta: 436795 10
muc6a2.dta : 904003 15
muc6b1.dta: 350823 10
muc6b2.dta : 29530 104
muc6b34.dta : 29530 30
muc7.dta: 250928 13
muc7ho.dta : 29532 9
muc8.dta : 29532 120
muc9.dta : 29532 49
tongcong.dta : 29532 51

Questionnaire

Household Questionnaire for Income & Expenditure Survey in English includes 11 Excel files.							
File name							
Cover.xls							
Sec01xls to Sec09.xls							
Balance.xls							
Note: Vietnamese version was also provided.							

Data dictionary

HouseholdDic.doc	Describes var	iable n	ame, type	e, format and label for Section 1 to 9
Example	Section 14			
Enumpie		etorage	dignlay	بب
	variable name	type	format	variable label ^{4/}
	tinh	int	%8.0a	province↔
	xa	long	%10.0f	commune+'
	hoso	long	%8.0g	household number+
	matv	long	%12.0f	individual code+
	m1c2	byte	%8.0g	sex: 1=male, 2=female+
	mic3	byte	%8.0g	Relationship to head of household: 1=Head, e^{i} 2=wife/husband, 3=child, 4=child in law e^{i}
	m1c4 1	byte	%10.0g	date of birth (Solar calendar)?- 2 numerals+
	m1c4_2	byte	%10.0g	Month of birth (Solar calendar)? - 2 numerals
	m1c4_3	int	%10.0g	Year of birth (Solar calendar)? - 4 numerals4
	mlc5	int	%8.0g	Age - in round ages↔
	mlc6	byte	%8.0g	The current marital status: 1=Never married; 2=married; 3=widowed; 4=divorced; 54
	gui	byte	%8.0g	quarter (of year)↔
	phieu	float	%9.0g	The income and expenditure questionnaire $\!\!\!\!^{\rm el}$

Codebook

Codebook_2002.xls	Describes variable name, topics, scope, length, code and description in each
	sheet of dataset

ample	A	В	С	D	E	F	G
1 1	Exp	planatory docum	ents for the VHLSS 2002				
2							
3	Nar	me of the datase	ıt	muc1: List of household member	13 varia	bles	
4							
5	No	Variable name	Topics	Scope	Length	Code	Description
6		1 tinh	province	All	8	101-823	Province code - Numeric
7		2 xa	commune	All	8	1010103-8231317	Commune code -Numeric
8		3 hoso	Household number	All	8	1001-58105	Numeric
9		4 matv	Individual number	All	8	100101-5810501	Numeric
10	1 1	5 m1c2	Sex	All	8		
11	(1	Male
12	1					2	Pemale
13	1	6 m1c3	Relationship to head of household	All	8		
14	1					1	l Head
15	5					2	2 Wife/husband
16	6					3	6 Child
17						4	Child in law
18	5					5	Parents
15	j					6	Sister/brother
20	1					7	Grandfather/grandmother
2*						8	3 Grandchild

Province code

Code of provinces and regions 2002.xls	Includes the next two list;
	List of provinces in VHLSS 2002 (61 provinces)
	List of regions in VHLSS 2002 (8 regions)
	Note: Definition of regions was not included here.

[VHLSS 2004]

Outline of the survey;

Survey plan.pdf	Description of objectives, contents, methodology, direction and				
	implementation of VHLSS2004				
Guidance to using	1. Contents of the VHLSS2004				
database of	2. Survey sample of income and expenditure and representative level				
VHLSS2004.doc	3. Content and structure of VHLSS2004 database				
	4. Some items to pay attention to when using the VHLSS2004 database				

Data in STATA and SAS format

File names in STAT	'A format;				
[1] "ho1.dta"	"ho2.dta"	"m1_2_3a.dta"	"m10a_e3.dta"		
[5] "m10f.dta"	"m1b.dta"	"m3b.dta"	"m4a.dta"		
[9] "m4b11.dta"	"m4b12_14.dta	" "m4b15.dta"	"m4b16.dta"		
[13] "m4b21.dta"	"m4b22.dta"	"m4b31.dta"	"m4b32.dta"		
[17] "m4b41.dta"	"m4b42.dta"	"m4b51.dta"	"m4b52.dta"		
[21] "m4c1.dta"	"m4c2.dta"	"m4d.dta"	"m5a1.dta"		
[25] "m5a2.dta"	"m5b1.dta"	"m5b2.dta"	"m5b3_4.dta"		
[29] "m6.dta"	"m6a.dta"	"m6b.dta"	"m7.dta"		
[33] "m8.dta"	"m8_xdgn.dta"	"m91.dta"	"m92.dta"		
[37] "m93.dta"	"m941.dta"	"m942_44.dta	" "m9451.dta"		
[41] "m9452.dta"	"m9453.dta"	"m946.dta"	"m9471.dta"		
[45] "m9472.dta"	"m948.dta"	"m949.dta"			
Number of records	and variables in ea	ach data file;			
ho1.dta : 9188 12	8				
ho2.dta : 9188 39					
m1_2_3a.dta: 40438 49					
m10a_e3.dta : 43	76 89				
m10f.dta: 416 15					
m1b.dta : 20209 1	4				
m3b.dta : 16777 17					
m4a.dta : 37368 45					
m4b11.dta : 11102 20					
m4b12_14.dta : 3	1444 14				
m4b15.dta : 8493	11				
m4b16.dta : 5416	1 12				

m4b21.dta : 18225 15
m4b22.dta: 11311 19
m4b31.dta : 301 10
m4b32.dta : 299 18
m4b41.dta : 4593 9
m4b42.dta : 2533 21
m4b51.dta : 4243 15
m4b52.dta : 3046 20
m4c1.dta : 4544 25
m4c2.dta : 29152 19
m4d.dta : 9188 20
m5a1.dta : 129993 11
m5a2.dta : 290143 16
m5b1.dta : 108735 11
m5b2.dta : 116575 9
m5b3_4.dta : 9188 22
m6.dta : 9188 68
m6a.dta : 16512 13
m6b.dta : 79235 12
m7.dta : 9188 52
m8.dta : 9188 42
m8_xdgn.dta : 5233 19
m91.dta : 35317 23
m92.dta: 1219 14
m93.dta: 805 14
m941.dta: 11095 11
m942_44.dta: 31444 10
m9451.dta: 25119 10
m9452.dta: 16003 15
m9453.dta : 3347 13
m946.dta: 12229 11
m9471.dta: 737 17
m9472.dta: 262 15
m948.dta : 6506 19
m949.dta : 1061 20
muc6b34.dta : 29530 30

```
muc7.dta : 250928 13
muc7ho.dta : 29532 9
muc8.dta : 29532 120
muc9.dta : 29532 49
tongcong.dta : 29532 51
```

Questionnaire

Household Questionnaire for Income & Expenditure Survey in English includes 12 Excel files.				
File name				
Cover.xls				
Section01xls to Section10.xls				
Indicators.xls				
Note: Vietnamese version was also provided.				

Codebook

Codebook_2004.xls	Describes variable name, topics, scope, length, code and description in each					
	sheet of data	set				
Example	AB	С	D	E	F	G
	1 Explanatory docu	ments for the VHLSS 20	04			
	3 Name of the data	se M1_2_3A			variables	49
	4					
	5 No Variable na	am Topics	Scope	Length	Code	Description
	6 1 TINH	Province	All	8	101-823	Numeric
	7 2 HUYEN	District	All	8	1-53	Numeric
	8 3 XA	Commune	All	8	1-95	Numeric
	9 4 DIABAN	Enumeration area	All	8	1-105	Numeric
	10 5 HOSO	Household code	All	8	1-27	Numeric
	11 6 KY	Surveyed period	All	8	1-2	Numeric
	12 7 MATV	ID code	All	8	1-20	Numeric
	13 8 M1AC1	Family name of HH member	All	16		Character
	14 9 M1AC1A	Sure name of HH member	All	8		Character
	15 10 M1AC2	Sex of HH member	All	8		
	16				1	Male
	17				2	Female
	18 11 M1AC3	Relationship with HH head	All	8		
	19				1	Head
	20				2	Wife/husband
	21				3	Child
	22				4	Parents
					-	· · · · ·

Province code

Province code.xls	Includes the next list;
	List of provinces in VHLSS 2004 (64 provinces)
Remarks: Three provinces increased.	

Vietnam VHLSS Manual - Overall and survey process (Version 1.1)

[VHLSS 2006]

Outline of the survey;

Survey plan.pdf	Description	of	objectives,	contents,	methodology,	direction	and
	implementation of VHLSS2006						

Data in STATA and SAS format

File names in STATA fo	ormat;		
[1] "muc1a.dta"	"muc1b.dta"	"muc2a.dta"	"muc2b.dta"
[5] "muc2c.dta"	"muc2d.dta"	"muc2e.dta"	"muc3a1.dta"
[9] "muc3a2.dta"	"muc3b.dta"	"muc3c.dta"	"muc3d.dta"
[13] "muc3e.dta"	"muc3f.dta"	"muc3g.dta"	"muc3h.dta"
[17] "muc3i.dta"	"muc4a.dta"	"muc4b0.dta"	"muc4b11.dta"
[21] "muc4b12.dta"	"muc4b13.dta"	"muc4b14.dta"	"muc4b15.dta"
[25] "muc4b16.dta"	"muc4b161.dta"	"muc4b21.dta"	"muc4b22.dta"
[29] "muc4b31.dta"	"muc4b32.dta"	"muc4b41.dta"	"muc4b42.dta"
[33] "muc4b51.dta"	"muc4b52.dta"	"muc4c.dta"	"muc4c2.dta"
[37] "muc4d.dta"	"muc5a1.dta"	"muc5a2.dta"	"muc5b1.dta"
[41] "muc5b2.dta"	"muc5b3_4.dta"	"muc6.dta"	"muc6a.dta"
[45] "muc6b.dta"	"muc7.dta"	"muc8.dta"	"muc8_vayvon.dta"
[49] "ttchung.dta"			
Number of records and	variables in each dat	a file;	
muc1a.dta : 39071 17			
muc1b.dta : 18792 16			
muc2a.dta : 39071 34			
muc2b.dta : 39071 21			
muc2c.dta : 39071 31			
muc2d.dta : 39071 19			
muc2e.dta : 39071 16			
muc3a1.dta : 39071 12	2		
muc3a2.dta : 18524 1	8		
muc3b.dta : 39071 93			
muc3c.dta: 39071 17			
muc3d.dta : 6647 15			
muc3e.dta : 33826 11			
muc3f.dta : 39071 28			
muc3g.dta : 11058 28			

muc3h.dta : 9214 31
muc3i.dta : 21622 16
muc4a.dta : 39071 52
muc4b0.dta : 24080 16
muc4b11.dta : 14961 19
muc4b12.dta : 15499 12
muc4b13.dta : 3932 13
muc4b14.dta : 9791 13
muc4b15.dta : 8379 10
muc4b16.dta : 52513 11
muc4b161.dta : 16592 11
muc4b21.dta : 16577 14
muc4b22.dta : 10222 18
muc4b31.dta : 279 9
muc4b32.dta : 277 17
muc4b41.dta : 3932 13
muc4b42.dta : 2283 20
muc4b51.dta : 3653 14
muc4b52.dta : 2706 19
muc4c.dta: 4379 33
muc4c2.dta: 29333 9
muc4d.dta : 9189 22
muc5a1.dta : 131726 10
muc5a2.dta : 288564 15
muc5b1.dta : 114114 10
muc5b2.dta : 114952 8
muc5b3_4.dta : 9189 23
muc6.dta : 9189 67
muc6a.dta : 15897 12
muc6b.dta : 85890 11
muc7.dta : 9189 51
muc8.dta : 9189 37
muc8_vayvon.dta: 4987 20
ttchung.dta : 9189 141

Questionnaire

Household Questionnaire for Income & Expenditure Survey in English				
File name	VHLSS06_questionnaire.pdf (85 pages)			

Codebook

Codebook_2006.xls	Describes variable name, topics, scope, length, code and description in each					
	sheet of datase	t				
Example	A B 1 Explanatory document	C ts for the VHLSS 2006	D	E	F	G
1	2					
	3 Name of the dataset	MUC1A			variables	17
	5 No. Variable name ir	Topics	Scone	Length	Code	Description
	6 1 TINH	Province	All	Lengen	3 101-823	Character
	7 2 HUYEN	District	All	2	2 01-53	Character
	8 3 XA	Commune	All	2	2 01-95	Character
	9 4 DIABAN	Enumerator area	All	3	3 001-105	Character
	10 5 HOSO	Household code	All	8	3 13-25	Numeric
	11 6 MATV	ID code	All	8	3 1-17	Numeric
	12 7 M1AC2	Sex of HH member	All	8	3	
	13				1	Yes
	14				2	No
	15 8 M1AC3	Relationship with HH member	All	8	3	
	16				1	Head
	17				2	Spouse
	18				3	Children
	19				4	Parents
	20				5	Grandfather/Grandmother
	21				6	Grandchild
	22				7	Other relation
	23				9	Missing

Province code

Province code.xls	Includes the next list;			
	List of provinces in VHLSS 2006 (64 provinces)			
Note: Not confirmed whether this is the same as VHLSS 2004.				

Survey results

Survey report of	Includes the results of VHLSS 2002, 2004 and 2006 as time series				
VHLSS 2006					
File name	Introduction.pdf and Part 01.pdf to Part 11.pdf				
Content of the report	Result of the Vietnam household living standards survey 2006				
	<image/> <section-header><section-header><section-header></section-header></section-header></section-header>	 Linving standard of the population reflected through the household living standard survey 2006 Section 1. Demography Section 2. Education Section 3. Labour - Employment Section 4. Health and Health care Section 5. Income Section 6. Consumption expenditure and consumption enxenditure for living Section 7. Fixed assets and durable goods Section 8. Housing, electricity, access to safe dringking water, sanitary and internet Section 9. Involvement in poverty alleviation programmes Section 10. Household businesses Section 11. Characteristics of communes 			
Note	The survey report is also a	vailable at NSO's website.			
	http://www.gso.gov.vn/default_en.aspx?tabid=515&ItemID=8183				
	(Accessed on 11 July 2014)				

[Weight data]

Weight is unique within xa (commune), which is psu. The following weight data by xa was provided;

File			De	Description			No of records	
weight_by_xa02.dta				Weight for VHLSS 2002 by psu			2,901	
weight_by_xa04.dta				Weight for VHLSS 2004 by psu			3,061	
weight_by_xa06.dta			a We	Weight for VHLSS 2006 by psu			3,063	
F	Example of VHLSS 2006:							
ſ	R データエディタ							
	tinh huy		huyen	xa	wt9	var5	I	
	1	101	1	3	3107.318			
	2	101	1	9	3092.521			
	3	101	1	15	3085.123			
	4	101	1	17	3099.92			
	5	101	1	21	3085.123			
	6	101	1	23	3099.92		T	
	7	101	3	3	2988.944		T	
	8	101	3	11	3003.741			

3.2 Other sources of metadata

□ IHSN (International Household Survey Network)

VHLSS 2002 (http://catalog.ihsn.org/index.php/catalog/3284/related_materials)

(Accessed on 11 July 2014)

The following documents are available;

Documentation

Download the questionnaires, technical documents and reports that describe the survey process and the key results for this study.

Questionnaires	
- Ancoronautica	
Household Living Standards Survey 2001-2002: Household Questionnaire	🔑 575.73 КВ
Vietnam Household Living Standards Survey 2002: Commune Questionnaire	🔑 133.07 КВ
Technical Documents	
Vietnam Household Living Standards Survey (VHLSS), 2002 and 2004: Basic Information	🔎 202.93 КВ
General Introduction of Vietnam Household Living Standards Survey (VHLSS)	🔑 121.6 КВ
Main difference between VLSS 1992/1993, VLSS 1997/1998 and VHLSS 2001/2002	🔑 77.39 КВ
Codes of Provinces and Regions	J 52 19 VP

Document	Description		
VHLSS 2002 and 2004	This document was prepared by the officials from Social & Environment		
Basic Information	Statistics Department, Vietnam General Statistics Office. It describes on		
(47 pages)	sample design of VHLSS 2002 and 2004 in detail.		
Main difference	VHLSS 2002 is a development of VLSS conducted during 1990's.		
between VLSS	It describes on the main difference between VLSS and VHLSS.		
1992/1993, VLSS			
1997/1998 and VHLSS			
2001/2002 (5 pages)			

VHLSS 2004 (http://catalog.ihsn.org/index.php/catalog/3285/related_materials)

The following documents are available;

Documentation

Download the questionnaires, technical documents and reports that describe the survey process and the key results for this study.

ocument	Description			
Stata programs		鬼 1.72 КВ		
Other Materials				
VHLSS 2004: Explanation of the Data Files				
Vietnam Household Living Standards Survey (VHLSS) 2004: Operational Handbook				
Vietnam Household Living Standards Survey (VHLSS), 2002 and 2004: Basic Information				
Technical Documents				
Vietnam Household Living Standards Survey 2004: Commune Questionnaire				
🔛 Household Living Standards Survey 2004: Questionnaire on Household Survey (Vietnamese)				
Household Living Standards Survey 2004: Questionnaire on Household Survey				

Document	Description
VHLSS 2004	It is the operational handbook in the field. It describes on the tasks and
Operational Handbook	responsibilities of team leader, enumerator and supervisor, as well as how
(124 pages)	to fill in questionnaire in detail.

□ World bank: Living Standards Measurement Study

(http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,c ontentMDK:21709814~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html Accessed on 07 August 2014)

The World Bank				Home • Site Ma	p • Index • FAQs	• Contact Us	
About Countries	Data & Research	Learning	News	Projects & Operations	Publications	Topics	
DEVELOPMENT Living Standards Search LSMS V GO RESEARCH Measurement Study						GO	
Home > Data & Resea	Home > Data & Research > Research > LSMS > Viet Nam 2004 🖂 Email 🎒 Print						
About LSMS	Viet Nam 20	04					
 LSMS - Integrated Surveys on Agricul 	Household	Household Living Standards Survey					
🗄 Datasets	Documentat	ion 🗌	Data				
	sample Size:		9	188			
E Survey Design & Analysis Tools Analysis Tools Analysis Tools Survey Design & Analysis Tools Analysis Tools Survey Design & Analysis Tools Survey Design & Analysis Tools Survey Design & Survey Desig	Coverage: Domains:		NU	National Urban/rural; Regions (Hanoi, Other cities, Other urban ar			
 Methodological Experiments 	Collection Dates:			Northern Nountains, Red River Delta, North Central, Central Coast, Central Highlands, Southeast, Mekong Delta) May and September 2004			
 Comparative Living Standards Project (CLSP) 	Panel: Executing Agency Access Policy:	r.	Ye <u>G</u> In	Yes (partial) with 2002 and 2006 General Statistical Office Information on how to obtain the data			
🛨 Reports & Researc	Comments:	veys for this c	Juntry: <u>1992-93, 1997-98, 2002,</u> 2006				
🗄 Training							
 Archived Correction and Modifications 	ins						
🗄 Core Team							
Site Tools							
Contact Us							
		Documentation File Name Description and Format Basic Describes the design of the survey and its coverage for potential users and provides general information about the 2002 and 2004 VHLSS.					
	File Name						
	Basic Information						
"Survey design and sample design in household budget surveys" by Hans Pettersson, Statistics Sweden (14 pages) (http://unstats.un.org/unsd/hhsurveys/pdf/Chapter_24.pdf) (Accessed on 11 July 2014)

This paper addresses some issues on survey design and sample design for household budget surveys. It explained the <u>square root allocation method</u> as one compromise among conflicting demands from National Accounts, and government planners and policy analysts. Square root allocation has been used for VHLSS. He had designed the sampling for VHLSS.

□ Enumerators' manual of VHLSS 2004

"Operational Handbook: VHLSS 2004" by GSO, April 2004, 124 pages (Available at IHSN website. Contents copying and changing documents are restricted.)

Below is the contents of the manual. Especially, "Section 1. Part A. List of household members" is very important.

CONTENTS

Part I: General principles	
Part II: Questionnaire for interviewing households10	
Section 0: Information on the cover of the household questionnaire	
Section 1. Part A. List of household members	
Section 1. Part B. Household members in VHLSS 200224	
Section 2. Education, training and vocational training	
Section 3. Health care	
Section 4. Income	
Section 5. Expenditure	
Section 6. Fixed assets and consumer durables	
Section 7. Housing	
Section 8. Participation in poverty reduction programmes	
Section 9. Agriculture, forestry and fisheries (expanded)	
Section 10. Business and production trades, non-agriculture, forestry and fisheries service (expanded)	s
Part III: Questionnaire for interviewing communes	

4. Map of Provinces and Regions (as of 2006)

Map of Provinces

The below map of Vietnam exhibiting its 59 provinces and 5 centrally controlled municipalities at the time of VHLSS 2004 and 2006. Ha Tay province, located in the south west of "2 Ha Noi" was merged with Ha Noi municipality in 2008. The number of provinces is 63 in 2014.



Map of regions

Provinces are grouped into eight regions.

Note: From VHLSS 2010, Vietnam is divided into 6 regions in GSO's publication.

Code	Region	Province codes	Note: Regions from 2001		
1	Red River Delta	101-117	Red River Delta		
2	North East	201-225	Northern Midland and		
3	North West	301-305	Mountains		
4	North Central	401-411	North Central and Central		
5	South Central Coast	501-511	Coast		
6	Central Highlands	601-607	Central Highlands		
7	South East	701-717	South East		
8	Mekong River Delta	801-823	Mekong River Delta		



5. Panel Data (Trial)

It is an advantage of survey design of VHLSS to make use of panel data.

What is panel data?

The term panel data refers to observations obtained over multiple time periods for the same households or individuals.

{Xijt} i: household or individual, j: variable, t: year

In statistical surveys, panel data is essentially a set of pairs of household identifiers (ID06, ID04) and pairs of individual identifiers (PID06, PID04). Each identifier is uniquely linked with household or individual data in VHLSS 2006 and 2004.



ID: Household identifier PID: Individual identifier 06: VHLSS 2006 04: VHLSS 2004

How can the design of VHLSS make panel data?

VHLSS 2006 has the question on whether the household was surveyed in VHLSS 2004 in Section 1b. And if yes, household identification of VHHLSS 2004 was recorded by the survey team leader based on the sample list of VHLSS 2004. Finally the data set ttchung has the variable of "ghepho" on whether matched with VHLSS 2004.

As for matching household members, the data set MUC1B, list of household members at the time of

40

VHLSS 2004 was prepared. Basic information was recorded by the survey team leader based on the list of household members of VHLSS 2004. It has the variables of person number in 2004, sex in 2004 and age in 2004 in addition to name in 2004, as well as person number in VHLSS 2006 if a person is the household member at the time of VHLSS 2006. And if a person is no longer living in the household at the time of VHLSS 2006, the reason was asked. Finally, the data set has the variable of ghep on whether matched with VHLSS 2004.

VHLSS 2006	VHLSS 2004
♦list of sample household a half surveyed in 2004 pairs (ID06, ID04)	♦list of sample households →questions on household
♦list of household members (PID06)	♦list of household members→questions on household members
 ♦ list of household members surveyed in 2004 If a person is a member in 2006, YES: pairs (PID06, PID04) No: Move-out/died info. 	
Û	$\hat{\mathbf{U}}$
Micro data sets in 2006	Micro data sets in 2004

What is available from panel data?

• Household panel data show the transition of household-level variables over time. It also provides information on the household which was selected for VHLSS 2006 but had removed to other EA.

Transition of households

VHLSS 2004

VHLSS 2006



- For instance, as described in the survey report, GSO makes use of panel data to compare directly household income per capita at household level.
- Individual panel data provide information on changes of individual-level variables of persons who were surveyed both in 2004 and 2006, that is, remained in the same households during 2004 and 2006.

In addition, it reveals persons who were surveyed in 2004 but have moved out of the household or died, as well as persons who moved in the household or were born during 2004 to 2006.

Transition of household members



Possible errors during conducting surveys

If the pairs of (ID06, ID04) and (PID06, PID04) are perfect, making panel data is a task of merging data sets between 2006 and 2004. However, it is natural to consider errors in the pairs of (ID06, ID04) and (PID06, PID04) during the survey process.

Possible errors are as the next;

- Recording errors when a survey team leader filled in questionnaire from the lists of VHLSS 2004.
- Data entry errors by an operator.
- Response errors by a sample household

Therefore, we will treat pairs of (ID06, ID04) and (PID06, PID04) available in data sets as candidates of panel data.

Strategy for creating panel data

- 1) To prepare data sets and variables for matching
- 2) To verify the information for matching
- 3) To drop the unqualified records from candidates of panel
- 4) To create panel at household level and individual level

Preparation for making panel data

Goal: The below data sets with household and individual identifiers will be prepared.

Data set	Original	Description	Household		Individual	
			identifier		identifier	
HLD06	TTCHUNG	All households surveyed in	ID06 ID04			
		2004 and link data				
HLD04	HO1	All households		ID04		
LIN06	MUC1B	Individual-level link data	ID06	ID04	PID06	PID04
MEM06	MUC1A	All household members	ID06		PID06	
		belonging to households				
		surveyed in 2004				
MEM04	M1_2_3A	All household members		ID04		PID04

VHLSS 2006

List of data sets and variables related to panel (VHLSS 2006)

Data set	Variable	Description
TTCHUNG		Summary file for all households (9,189)
	m1c1	Surveyed in 2004? (Yes: 4,298 No: 4,891)
	ghepho	Matched with VHLSS 2004?
		(0: 4,891 Not surveyed in 2004 1: 4,267 Data in 2004 9: 31)
	tinh04	
	huyen04	
	xa04	
	diaban04	

	ttnt04	
	hoso04	
MUC1A		All household members (39,071)
MUC1B		Household members surveyed in VHLSS 2004 (18,792)
	m1bc3	Person number in 2004
	m1bc4	Sex in 2004
	m1bc5	Age in 2004
	ghep	Matched with VHLSS 2004? (0: 671 Not surveyed in 2004 1:
		17,039 Data in 2004 9: 37 NA: 1,045)
	m1bc6	Household member in 2006? (1: 17,076 Yes 2: 1,716 No)
	m1bc7	Person number in 2006
	m1bc8	Reason for no longer living in the household
	m1bc9	Reason for moving out of the household
	m1bc10	Province code work in

Generating household-level data frame <u>HLD06</u> from TTCHUNG

```
> d<-subset(ttchung,m1c1==1)
```

```
> dim(d)
```

```
[1] 4298 144
```

Renamed the household identifier ID as ID06

```
>d["ID06"]<-d$ID
```

✓ Generated the household identifier ID04.

- $> d["ID04"] <- as.character(d\$hoso04 + (10^{2})*d\$diaban04 + (10^{5})*d\$xa04 + (10^{2})*d\$xa04 + (10^{2})*axa04 + (10^{2})*axa04 + (10^{2})*d\$xa04 + (10^{2})*dxa04 + ($
- $+ (10^{7})*d$ huyen 04 + (10^9)*d tinh 04)
- > HLD06<-d

Remarks: Out of 4,298 households, 3,667 households have the same ID both in 2004 and 2006.

```
> table(HLD06$ID==HLD06$ID04,useNA="ifany")
FALSE TRUE <NA>
631 3667 4891
```

Generating individual-level data frame MEM06 from MUC1A

> d<-muc1a	
> str(d[1:6])	
'data.frame':	39071 obs. of 6 variables:
\$ tinh : chr	"101" "101" "101" "101"
\$ huyen : chr	"01" "01" "01" "01"
\$ xa : chr	"03" "03" "03" "03"
\$ diaban: chr	"014" "014" "014" "014"
\$ hoso : int	15 15 15 15 19 19 19 19 24 24
\$ matv : int	1 2 3 4 1 2 3 4 1 2

✓ Generated household identifier ID06 in MUC1A.

 $> d["ID06"] <-as.character(d\$hoso+(10^2)*as.integer(d\$diaban)+(10^5)*as.integer(d\$xa)+(10^2)*as.integer(d\xa)+(10^2)*as.inte$

+ (10^7)*as.integer(d\$huyen)+(10^9)*as.integer(d\$tinh))

✓ Generated individual identifier PID06 as a combination of ID06 and person number (matv).

> d["PID06"]<-paste(d\$ID06,formatC(d\$matv,width=2,flag="0"),sep="")

> MEM06<-subset(d,is.element(d\$ID06,HLD06\$ID06))

> dim(MEM06)

[1] 18419 23

Generating individual-level data frame LIN06 from MUC1B

> d < -muc 1b

Renamed the household identifier ID as ID06

> d["ID06"] <-d\$ID

 ✓ Generated individual identifier PID06 as a combination of ID06 and person number in 2006 (m1bc7) if the person was a household member in 2006.

>t<-table(d\$m1bc6)

```
> names(t)<-c("Member in 2006","No")</pre>
```

> t

Member in 2006	No
17076	1716

> d["PID06"] <-ifelse(d\$m1bc6==1,paste(d\$ID06,formatC(d\$m1bc7,width=2,flag="0"),sep=""),NA)

- ✓ Appended ID04 by merging with HLD06, and generated PID04 as a combination of ID04 and person number in 2004 (m1bc3).
- > d<-merge(d,HLD06[c("ID06","ID04")],key="ID06",all.x=T)
- > d["PID04"]<-paste(d\$ID04,formatC(d\$m1bc3,width=2,flag="0"),sep="")

>LIN06<-d

✓ LIN06 consists of 18,876 household members at the time of VHLSS 2004 in sample households surveyed in VHLSS 2004, of which 17,076 persons are household members in VHLSS 2006, that is, they have PID06.

```
> table(is.na(LIN06$PID04),useNA="ifany")
FALSE
18792
> table(is.na(LIN06$PID06),useNA="ifany")
FALSE TRUE
17076 1716
```

VHLSS 2004

Data set	Variable	Description		
HO1		Summary file for all households (9,188)		
	m1c1	Surveyed in 2002? (1 Yes: 4,476 2 No: 4,712)		
	tinh02			
	xa02			
	ttnt02			
	hoso02			
	quy02	Quarter code in 2002		

List of data sets and variables related to panel (VHLSS 2004)

M1_2_3A		All household members (40,438)
M1B		Household members surveyed in VHLSS 2002 (20,209)
	m1bc3	Person number in 2002
	m1bc4	Sex in 2002
	m1bc5	Age in 2002
	m1bc6	Household member in 2004? (1: 18,353 Yes 2: 1,856 No)
	m1bc7	Person number in 2004
	m1bc8	Reason for no longer living in the household
	m1bc9	Reason for moving out of the household

Generating household-level data frame HLD04 from HO1

> d<-ho1

Generated household identifier ID04.

 $> d["ID04"] <- as.character(d\$hoso+(10^{2})*d\$diaban+(10^{5})*d\$xa+(10^{7})*d\$huyen+(10^{9})*d\$tinh)$

Generating individual-level data frame <u>MEM04</u> from M1_2_3A

> d<-m1_2_3a

✓ Generated household identifier ID04 in $M1_2_3A$.

 $> d["ID04"] <- as.character(d\$hoso+(10^{2})*d\$diaban+(10^{5})*d\$xa+(10^{7})*d\$huyen+(10^{9})*d\$tinh)$

- ✓ Generated individual identifier PID04 as a combination of ID04 and person number (matv).
- > d["PID04"]<-paste(d\$ID04,formatC(d\$matv,width=2,flag="0"),sep="")

> MEM04<-d

Verifying data sets and variables for matching

✓ To generate data set hp1 as pairs (ID06, ID04) from HLD06, which is a candidate for household-level panel.

✓ To state two conditions for hp1 and seven conditions for LIN06 essential for panel, and prepare flags on whether conditions are satisfied or not for each record in the datasets.

Tow conditions for hp1

- 1) ID04 in hp1 should be in HLD04.
- 2) This linkage should be one to one relationship. There should be no duplicated ID04 in hp1.

Condition 1: ID04 in hp1 should be in HLD04.

- ✓ Out of 4,298 households surveyed in VHLSS 2004, the number of households linked with HLD04 is 4,219.
- # Appended flag1 for unmatched ID04 in hp1
- > hp1["flag1"]<-ifelse(is.element(hp1\$ID04, HLD04\$ID04), 1, 0)

```
> table(hp1$flag1)
```

```
0 1
```

```
79 4219
```

Example of errors > head(subset(hp1, flag1==0)) ID06 IDO4 flag1 105 101082500115 101082500115 0 727 105213700813 105213700813 0 742 105231900313 105231900313 0 743 105231900314 105231900314 0 744 105231900315 105231800318 0 1165 109051100413 109051100413 0

Condition 2: This linkage should be one to one relationship. There should be no duplicated IDO4 in hp1.

 \checkmark There are 13 duplicated ID04 in hp1 as follows.

```
> sum(duplicated(subset(hp1,flag1==1)$ID04))
[1] 13
```

List of duplicated ID04

```
> dup. ID04<-hp1[duplicated(hp1$ID04),"ID04"]
```

> dup. ID04

```
[1] "101093301019" "209214100214" "211033700214" "211073300814"
```

[5] "211095700513" "211114700114" "217012100813" "221072300615"

```
[9] "305101500215" "715130300514" "813070501613" "813191301219"
```

[13] "819070700115"

```
# Appended flag2 for duplicated records in hp1.
```

```
> hp1["flag2"]<-ifelse(!is.element(hp1$ID04, dup. ID04), 1, 0)
```

```
> t<-addmargins(table(hp1flag1, hp1flag2))[c(3, 2, 1), c(3, 2, 1)]
```

> rownames(t) <-c("Surveyed in 2004","Linked with HLD04","No")

```
> colnames(t)<-c("Total", "No duplication", "Duplicated")
```

```
> t
```

Total	No	duplication	Duplicated
-------	----	-------------	------------

Surveyed in 2004	4298	4272	26
Linked with HLDO4	4219	4193	26
No	79	79	0

✓ There are two different ID06 for each duplicated ID04. While we assume that one of the two same ID04 might be right and the other might be a result of data entry miss, removed the above 26 records tentatively because we cannot discuss on which is the right linkage at this moment.

Vietnam VHLSS Manual - Overall and survey process (Version 1.1)

✓ Created household-level panel hp2 consisted of 4,193 pairs of ID06 and ID04.

> hp2<-subset(hp1, flag1==1&flag2==1) [c("ID06", "ID04")]

Seven conditions for LIN06 in order to link the identical households

- 3) A person with PID04 should be in MEM04.
- 4) There should be no duplicated PID04 in LIN06.
- 5) The person's sex and age in LIN06 and MEM04 should be identical.
- 6) ID06 in LIN06 should be the same as ID06 in hp.
- 7) If a person in LIN06 was a household member in VHLSS 2006, that is, PID06 was defined in LIN06, the person with the PID06 should be in MEM06.
- 8) There should be no duplicated PID06 in LIN06.
- 9) The person's sex and age in LIN06 and MEM06 should be identical.

Condition 3: A person with PID04 in LIN06 should be in MEM04.

- \checkmark Out of 18,792 persons in LIN06, the number of those linked with records in MEM04 is 18,348.
- # Appended **flag3** for unmatched PID04 in LIN06
- > LIN06["flag3"]<-ifelse(is.element(LIN06\$PID04, MEM04\$PID04), 1, 0)
- > t<-addmargins(table(LINO6m1bc6, LINO6f1ag3))[c(3, 1, 2), c(3, 2, 1)]
- > rownames(t)<-c("Person in LIN04","Member in 2006","No (move-out/died)")
- > colnames(t)<-c("Total", "Linked with MEMO4", "No")

> t

	Total	Linked	with	MEM04	No
Person in LINO4	18792			18348	444
Member in 2006	17076			16685	391
No (move-out/died)	1716			1663	53

Condition 4: There should be no duplicated PID04 in LIN06.

✓ Among 18,348 PID04 in LIN06 which linked with MEM04, there are 42 duplicated PID04 as follows. Which record is right one will be discussed in the next condition 5.

```
> sum(duplicated(subset(LIN06,flag3==1)$PID04)==T)
[1] 42
```

```
# List of duplicated ID04
> dup.PID04<- subset(LIN06,flag3==1) [duplicated(subset(LIN06,flag3==1)$PID04)==T, "PID04"]
> length(dup.PID04)
[1] 42
```

```
# Appended flag4 for duplicated records in LINO6
```

```
> LIN06["flag4"]<-ifelse(!is.element(LIN06$PID04, dup.PID04), 1, 0)
```

```
> t<-addmargins(table(LINO6$flag3,LINO6$flag4))[c(3, 2, 1), c(3, 2, 1)]
```

```
> rownames(t)<- c("Person in LIN06", "Linked with MEMO4", "No")
```

> colnames(t)<-c("Total", "No duplication", "Duplicated")

```
> t
```

	Total	No duplication	Duplicated
Person in LINO6	18792	18708	84
Linked with MEMO4	18348	18264	84
No	444	444	0

```
# Examples of duplicated records
```

```
> for(j in dup.PID04) print(LIN06[LIN06$PID04==j,c("PID06","PID04")])
```

```
    PID06
    PID04

    178
    10109330101502
    10109330101902

    182
    10109330101902
    10109330101902

    PID06
    PID04
```

```
179 10109330101501 10109330101901

183 10109330101901 10109330101901

:

PID06 PID04

18011 81907070011303 81907070011503

18027 81907070011503 81907070011503
```

Condition 5: The person's sex and age in LINO6 and MEMO4 should be identical.

```
> d<-subset(LIN06,flag3==1)
```

8

```
> df1<-d[c("PID06", "ID06", "PID04", "ID04", "m1bc4", "m1bc5", "flag3", "flag4")]
```

> dim(df1)

[1] 18348

```
> colnames(df1)[5:6]<-c("lsex","lage")
```

> head(df1)

	PID06	ID06	PID04	ID04	lsex	lage	flag3	flag4
1	10101030141501	101010301415	10101030141502	101010301415	2	46	1	1
2	10101030141503	101010301415	10101030141504	101010301415	1	24	1	1
3	10101030141502	101010301415	10101030141503	101010301415	1	26	1	1
4	10101030141504	101010301415	10101030141505	101010301415	1	18	1	1
5	<na></na>	101010301415	10101030141501	101010301415	1	57	1	1
6	10101030141902	101010301419	10101030141902	101010301419	1	24	1	1

```
> df2<-MEMO4[c("PIDO4", "IDO4", "m1ac2", "m1ac5")]
```

> dim(df2)

[1] 40438 4

```
> colnames(df2)<-c("PID04", "ID04", "sex04", "age04")
```

> head(df2)

PID04 ID04 sex04 age04

```
1 10101030141301 101010301413
                                 2
                                     43
2 10101030141302 101010301413
                                     50
                                1
3 10101030141303 101010301413
                                     20
                                1
4 10101030141304 101010301413
                                     17
                                1
5 10101030141501 101010301415
                                     50″
                              1
6 10101030141502 101010301415
                                2
                                     46
```

53

```
> cond5<-merge(df1[c("PID06", "PID04", "|sex", "lage", "flag3", "flag4")],
```

```
+ df2[c("PID04", "sex04", "age04")], by="PID04", all.x=T)
```

- > dim(cond5)
- [1] 18348 8
- > head (cond5)

	PID04	PID06	lsex	lage	flag3	flag4	sex04	age04
1	10101030141501	<na></na>	1	57	1	1	1	50
2	10101030141502	10101030141501	2	46	1	1	2	46
3	10101030141503	10101030141502	1	26	1	1	1	26
4	10101030141504	10101030141503	1	24	1	1	1	24
5	10101030141505	10101030141504	1	18	1	1	1	18
6	10101030141901	10101030141901	2	52	1	1	2	52

 \checkmark Out of 18,348, the number of records satisfying the condition 5 is 17,723.

```
# Appended flag5 for inconsistent sex and age
```

- > cond5["flag5"]<-ifelse(cond5\$lsex==cond5\$sex04 &</pre>
- + abs (cond5\$lage-cond5\$age04) <=1, 1, 0)
- > table(cond5\$flag5)
 - 0 1

625 17723

Renamed cond5 as ip1 and generated ID06 and ID04 from PID06 and PID04

- > ip1<-cond5
- > ip1["ID06"]<-substr(ip1\$PID06, 1, 12)
- > ip1["ID04"]<-substr(ip1\$PID04, 1, 12)

Example of records with inconsistent sex and age between LINO6 and MEMO4
> head(ip1[ip1\$flag5==0, c(2, 1, 3, 4, 7, 8, 5, 6, 9)])

	PID06	PID04	lsex	lage	sex04	age04	flag3	flag4	flag5
1	<na></na>	10101030141501	1	57	1	50	1	1	0
59	<na></na>	10106070171305	2	21	2	23	1	1	0
125	10108030011401	10108030011401	2	44	2	42	1	1	0

129	10108030011501	10108030011501	1	46	1	44	1	1	0
174	10109330101501	10109330101901	1	42	2	49	1	0	0
176	10109330101502	10109330101902	2	41	2	18	1	0	0

Interpretation of errors:

- The pair of PID06=10109330101502 and PID04=10109330101902 is not proper.
- For PID04=10109330101902, there are two candidates, PID06=10109330101502 and PID06= 10109330101902 (condition 4).
- The pair of PID04=10109330101902 and PID06=10109330101902 is proper.
- ✓ Out of 84 records regarded as duplicated under the condition 4, that is, given flag4=0, a half of the 84 records are identified with the records with the same PID04 and the other half are not identified.

Therefore, data set ip2, a subset of records with flag5=1 in ip1 is candidates for panel at the moment.

> t<-addmargins(table(ip1\$flag4, ip1\$flag5))[c(3, 1, 2), c(3, 2, 1)]

```
> rownames(t)<-c("Linked with MEMO4","Duplicated","Not duplicated")
```

```
> colnames(t)<-c("Total", "Identified", "Not identified")</pre>
```

```
> t
```

	Total	Identified Not	identified
Linked with MEMO4	18348	17723	625
Duplicated	84	42	42
Not duplicated	18264	17681	583

```
> ip2<-subset(ip1,flag5==1)
```

> dim(ip2)

```
[1] 17723 11
```

✓ Again, applied the condition 4 for ip2 and found one duplication. We cannot distinguish from data of sex and age because they both satisfy the condition 5. We have to check household composition.

```
> sum(duplicated(ip2$PID04))
```

[1] 1

```
> ip2[ip2$PID04==ip2[duplicated(ip2$PID04), "PID04"], c("PID06", "ID06", "PID04", "ID04")]
```

Household composition of ID04=="819070700115"

> MEMO4[MEMO4\$ID04=="819070700115", c("PID04", "ID04", "m1ac2", "m1ac5")]

PID04 ID04 m1ac2 m1ac5

- 38860
 81907070011504
 819070700115
 2
 9

 38861
 81907070011503
 819070700115
 1
 11

 38862
 81907070011502
 819070700115
 2
 30
- 38863 81907070011501 819070700115 1 30

Household composition of ID06=="819070700113"

> MEM06[MEM06\$ID06=="819070700113", c ("PID06", "ID06", "m1ac2", "m1ac5")]

PID06 ID06 m1ac2 m1ac5 37561 81907070011301 819070700113 1 34

37562 81907070011302 819070700113 2 32

37563 81907070011303 819070700113 2 2

Household composition of ID06=="819070700115"

> MEMO6[MEMO6\$ID06=="819070700115", c("PID06", "ID06", "m1ac2", "m1ac5")]

PIDO6 IDO6 m1ac2 m1ac5

3757181907070011501819070700115132375728190707001150281907070011523237573819070700115038190707001151133757481907070011504819070700115211

✓ As a result of comparing household composition, the pair of PID06="81907070011502" and PID04="81907070011502" is right. The pair of PID06="81907070011302" and PID04="81907070011502" should be dropped from ip2.

> ip2[ip2\$PID04==ip2[duplicated(ip2\$PID04), "PID04"], c("PID06", "ID06", "PID04", "ID04")]

	ID04	PID04	ID06	PID06	
←Should be dropped	819070700115	81907070011502	819070700113	81907070011302	17598
←Right!	819070700115	81907070011502	819070700115	81907070011502	17599

Condition 6: ID06 in LIN06 should be the same as ID06 in hp.

✓ As for LIN06, a set of ID06 is the same as a set of ID06 in HLIN.
 > table(unique(LIN06\$ID06) ==HLIN\$ID06)
 TRUE
 4298

 \checkmark As for ip3, the number of unique ID06 is 4,167.

✓ Appended **flag6** to hp2.

> ip3. ID06<-unique(ip3\$ID06)

> length(ip3.ID06)

[1] 4167

> hp2["flag6"]<-ifelse(is.element(hp2\$ID06, ip3. ID06), 1, 0)

1

1

> head (hp2)

IDO4 flag6

```
1 101010301415 101010301415
```

ID06

```
2 101010301419 101010301419 1
```

```
16 101012301813 101012301813 1
```

```
31 101051100213 101051100213 1
```

```
32 101051100215 101051100215 1
```

```
33 101051100219 101051100219
```

> addmargins(table(hp2\$flag6))

0 1 Sum

40 4153 4193

 \checkmark Defined hp3 as a subset of hp2 satisfying the condition 6.

```
> hp3<-subset(hp2,flag6==1)
> dim(hp3)
[1] 4153 3
```

Condition 7: If a person in LINO6 was a household member in VHLSS 2006, that is, PIDO6 was defined in LINO6, the person with the PIDO6 should be in MEMO6. (Satisfied!)

 \checkmark If a person in LIN06 is a household member of VHLSS 2006, the person belongs to MEM06.

```
> t<-table(is.na(LINO6$PIDO6), !is.element(LINO6$PIDO6, MEMO6$PIDO6))
```

```
> rownames(t)<-c("Member in LIN06", "No")
```

```
> colnames(t)<-c("Belong to MEMO6","No")
```

> t

			Belong	to	MEM06	No
Member	in	LIN06			17076	0
No					0	1716

Condition 8: There should be no duplicated PID06 in LIN06. (Satisfied!)

```
> sum(duplicated(subset(LIN06,m1bc6==1)$PID06))
[1] 0
```

Condition 9: The person's sex and age in LINO6 and MEMO6 should be identical.

- 4 10101030141504 10101030141503 1 24

```
5 10101030141505 10101030141504
                                       18
                                   1
6 10101030141901 10101030141901
                                   2
                                       52
7 10101030141902 10101030141902
                                       24
                                   1
> df2<-MEM06[c("PID06", "m1ac2", "m1ac5")]
> colnames(df2)<-c("PID06", "sex06", "age06")
> dim(df2)
[1] 18419
              3
> head(df2)
           PID06 sex06 age06
1 10101030141501
                     2
                          48
2 10101030141502
                          28
                     1
3 10101030141503
                     1
                          26
4 10101030141504
                     1
                          20
5 10101030141901
                     2
                          54
6 10101030141902
                     1
                          26
> cond9<-merge(df1, df2, by="PID06", all.x=T)
> dim(cond9)
[1] 17722
              6
> head (cond9)
           PID06
                          PID04 |sex |age sex06 age06
1 10101030141501 10101030141502
                                   2
                                               2
                                        46
2 10101030141502 10101030141503
                                       26
                                               1
                                   1
3 10101030141503 10101030141504
                                       24
                                   1
                                               1
4 10101030141504 10101030141505
                                   1
                                       18
                                               1
5 10101030141901 10101030141901
                                   2
                                       52
                                               2
6 10101030141902 10101030141902
```

 \checkmark Out of 17,722, the number of records satisfying the condition 9 is 16,103.

1 24 48

28

26

20

54

26

1

Vietnam VHLSS Manual - Overall and survey process (Version 1.1)

- # Appended **flag9** for inconsistent sex and age
- > cond9["flag9"]<-ifelse(cond9\$lsex==cond9\$sex06 &</pre>
- + abs (cond9\$lage+2-cond9\$age06) <=2, 1, 0)
- > table(cond9\$flag9)

0 1

18 16102

List of not identified records in ip3

> subset(cond9, flag9==0)

	PID06	PID04	lsex	lage	sex06	age06	flag9
3644	20705090041302	20705090041302	2	38	2	43	0
3708	20709010061403	20709010061404	1	16	1	21	0
7990	41101490141403	41101490141405	2	36	1	38	0
8064	41107310111305	41107310111305	2	5	2	10	0
10125	60321020051303	60321020051303	2	11	1	13	0
10126	60321020051304	60321020051304	2	5	1	7	0
10972	70109030301402	70109030301402	1	42	1	47	0
14571	81303050081404	81303050081406	2	8	1	16	0
14687	81311010131504	81311010131504	1	21	2	22	0
14726	81313050161303	81313050161905	2	21	1	23	0
14929	81525030151901	81525030151901	1	47	1	59	0
14931	81525030151903	81525030151903	1	28	1	36	0
15029	81601090041501	81601090041502	2	46	2	52	0
15030	81601090041502	81601090041503	2	43	2	50	0
15091	81607130081301	81607130081301	1	36	1	43	0
15190	81611070141504	81611070141504	2	8	1	10	0
16064	82309150131501	82309150131501	1	41	1	46	0
16077	82311070171503	82311070171506	2	15	2	27	0

✓ Generated data frame ip4 consisted of pairs of PID06 and PID04 as a subset of cond9 with flag9=1. Each pair of PID06 and PID04 identifies uniquely records in MEM06 and MEM04.

> ip4<-subset(cond9,flag9==1)[c("PID06","PID04")]

> ip4["ID06"]<-substr(ip4\$PID06,1,12)

> ip4["ID04"]<-substr(ip4\$PID04,1,12)

> dim(ip4)

[1] 16102 4

> head(ip4)

	PID06	PID04	ID06	ID04
1	10101030141501	10101030141502	101010301415	101010301415

60

2	10101030141502	10101030141503	101010301415	101010301415
3	10101030141503	10101030141504	101010301415	101010301415
4	10101030141504	10101030141505	101010301415	101010301415
5	10101030141901	10101030141901	101010301419	101010301419
6	10101030141902	10101030141902	101010301419	101010301419

- ✓ Comparing household size of individual panel data in 2004 with HLD04, 2,821 households are inconsistent but in 1,344 households the number of records of panel data within the household is smaller than the household size of the same household surveyed in 2004.
- > t<-tapply(ip4\$PID04, ip4\$ID04, length)
- > ip4. hhsz<-data.frame(ID04=names(t), ip. hhsz=as.vector(t), row.names=NULL)
- > ip4. hhsz2<-merge(ip4. hhsz, HLD04[c("ID04", "tsnguoi")], by="ID04", all.x=T)
- > head(ip4.hhsz2)

IDO4 ip.hhsz tsnguoi

- 1 101010301415 4 5
- 2 101010301419 3 3
- 3 101012301813 4 4
- 4 101051100213 2 6
- 5 101051100215 6 6
- 6 101051100219 4 4
- > table(ip4.hhsz2\$ip.hhsz==ip4.hhsz2\$tsnguoi,useNA="ifany")
- FALSE TRUE
- 1344 2821

✓ Finally, the individual-level panel data consisted of 11,411 pairs of (PID06, PID04) were generated, which is consistent with the household size in 2004.

```
> ip4. hhsz["flag"]<-ifelse(ip4. hhsz2$ip. hhsz=ip4. hhsz2$tsnguoi, 1, 0)
```

```
> table(ip4.hhsz$flag)
```

0 1

```
1344 2821
```

- > ip5. ID04<-subset(ip4. hhsz, flag==1)\$ID04
- > length(ip5.ID04)
- [1] 2821

> ip5<-subset(ip4, is.element(ip4\$ID04, ip5.ID04))

> dim(ip5)

[1] 11411 4

> head(ip5)

PID06PID04ID06ID045101010301419011010103014190110101030141910101030141961010103014190210101030141902101010301419101010301419710101030141904101010301419031010103014191010103014198101012301813011010123018130110101230181310101230181391010123018130210101230181302101012301813101012301813101010123018130310101230181303101012301813101012301813

Example:

Panel data ID06=" 101010301419" and ID04=" 101010301419" Household members of ID04=="101010301419" in 2004 and 2006 MEM04:

```
> d<-MEM04[MEM04$ID04=="101010301419", c(51, 10:15)]
> colnames(d)<-c("PID04", "sex", "relation", "month", "year", "age", "marital")
> d
```

PIDO4 sex relation month year age marital

10	10101030141901	2	1 1	1952	52	3
11	10101030141902	1	3 1	1980	24	1
12	10101030141903	1	3 2	1987	17	1

MEMO6:

```
> d1<-MEM06[MEM06$ID06=="101010301419", c(23, 7:12)]
```

> colnames(d1) <- c("PID06", "sex", "relation", "month", "year", "age", "marital")

> d1

PIDO6 sex relation month year age marital

5	10101030141901	2	1	1 1952	54	3
6	10101030141902	1	3	1 1980	26	2
7	10101030141903	2	3	11 1980	25	2
8	10101030141904	1	3	2 1987	19	1

✓ In the above case, the household consisted of a widowed mother and two sons of age 24 and 17 in 2004. The elder son got married during 2004 and 2006, and his wife was surveyed in 2006.

```
♦ Topic: Household size in 2004
```

Comparing LINO6 with HLDO4, the household sizes in 2004 of LINO6 are consistent in 4,066 households.

```
> t<-tapply(LIN06$PID04,LIN06$ID04,length)
```

- > ip. hhsz<-data. frame(ID04=names(t), ip. hhsz=as. vector(t), row. names=NULL)
- > ip.hhsz2<-merge(ip.hhsz,HLD04[c("ID04", "tsnguoi")], by="ID04", all.x=T)

```
> head(ip.hhsz2)
```

IDO4 ip.hhsz tsnguoi

```
1 101010301415
                    5
                            5
2 101010301419
                    3
                            3
3 101012301813
                    4
                            4
4 101051100213
                    6
                            6
5 101051100215
                    6
                            6
6 101051100219
                    4
                            4
> table(ip.hhsz2$ip.hhsz=ip.hhsz2$tsnguoi,useNA="ifany")
FALSE TRUE <NA>
  140 066
             79
```

 \diamond Topic 2: Difficulties of matching individual data

Case of ID04=="101010301415"

MEMO4:

```
> d<-MEM04 [MEM04$ID04=="101010301415", c (51, 10:15)]
```

```
> colnames(d) <-c("PIDO4", "sex", "relation", "month", "year", "age", "marital")
```

```
> d
```

PID04 sex relation month year age marital

5	10101030141501	1	1	5 1954	50	2
6	10101030141502	2	2	7 1957	46	2
7	10101030141503	1	3	11 1977	26	1
8	10101030141504	1	3	3 1980	24	1
9	10101030141505	1	3	3 1986	18	1

LIN06:

✓ In the above case, the age 57 of PID04=" 10101030141501" in LIN06 should be read as 50. It may be possible to revise errors manually in such way. However, how to make it programmable?

Example of usage at household level:

Changes of consumption expenditure per capita between 2004 and 2006

- # Household-level panel
- > hp5<-ip5[!duplicated(ip5\$ID06),c("ID06","ID04")]
- > HLD06["pcexp06"]<-HLD06\$chidsbq
- > HLD04["pcexp04"]<-HLD04\$chidsbq
- # Linked the pcexp data of HLD06 and HLD04
- > pcexp.panel<-merge(hp5,HLD06[c("ID06","pcexp06")])</pre>
- > pcexp.panel<-merge(pcexp.panel,HLD04[c("ID04","pcexp04")])
- > dim(pcexp.panel)

4

- [1] 2821
- # Generated the variable of increasing rate of pcexp between 2004 and 2006.
- > pcexp.panel["rate"]<-(pcexp.panel\$pcexp06/pcexp.panel\$pcexp04-1)*100</pre>
- > head(pcexp.panel)

	ID04	ID06	pcexp06	pcexp04	rate
1	101010301419	101010301419	1226	753	62. 81541
2	101012301813	101012301813	1111	814	36. 48649
3	101051100215	101051100215	430	373	15. 28150
4	101051100219	101051100219	942	494	90. 68826

5	101053301214	101053301214	778	554 40). 43321
6	101053301215	101053301215	968	700 38	3. 28571

Summary:

- ✓ Regarding VHLSS 2004 and 2006, household-level panel consisted of 4,193 household identifiers was generated.
- ✓ Also, individual-level panel consisted of 11,411 person identifier was generate.
- ✓ To make panel more complete, data entry and data check of identification items should be carried out more carefully.

6. Population and Household

6.1 Ethnicity

Kinh people are the majority ethnic group of Vietnam.
 Regarding household head's ethnicity, 88% of households are Kinh. The share of Kinh households is high especially in Red River Delta (99%), North Central (91%), South Central Coast (95%), South East (92%) and Mekong River Delta (93%).
 On the other, the share of Kinh is only 11% in rural of North West.

```
> d<-ttchung[c("ID", "wt", "tinh", "ttnt", "tsnguoi", "dantoc", "phdich")]
```

- # Generated the variable kinh: 1=Kinh, 2=Non-kinh
- > d["kinh"]<-ifelse(d\$dantoc==1, 1, 2)
- > table(d\$kinh)
 - 1 2
- 7745 1444
- # Weighted number of households by region and urban/rural
- > denominator<-addmargins(tapply(d\$wt,list(substr(d\$tinh,1,1),d\$ttnt),sum,na.rm=T))
- # Weighted number of Kinh households by region and urban/rural
- > d1<-subset(d,kinh==1)
- > numerator<-addmargins(tapply(d1\$wt, list(substr(d1\$tinh, 1, 1), d1\$ttnt), sum, na.rm=T))

Share of Kinh households by region and urban/rural

- > share<-round (numerator/denominator*100, 1) [c (9, 1:8), c (3, 1, 2)]
- > colnames(share)<-c("Total","Urban","Rural")
- > rownames(share)<-c("Vietnam", region.name)
- > share

	Total	Urban	Rural
Vietnam	87.8	94. 3	85.3
Red River Delta	99.4	100.0	99.2
North East	62.4	85.7	55.8
North West	24. 2	84. 6	11.1

North Central	91.2	98. 2	90.0
South Central Coast	94. 8	98.7	93.1
Central Highlands	71.1	86. 2	64.3
South East	92. 1	92.4	91. 6a
Mekong River Delta	93. 2	95.0	92.7

✓ Out of 1,444 non-kinh sample households, 381 households needed the interpretation service when conducting the survey.

It suggests that about one fourth non-Kinh households have communication problem with Kinh people.

> m<-tapply(d\$wt,list(d\$kinh,d\$phdich),length)

```
> m[is.na(m)]<-0
```

> m <- addmargins(m) [c(3, 1, 2), c(3, 1, 2)]

```
> rownames(m) <-c("Total", "Kinh", "Non-kinh")</pre>
```

```
> colnames(m)<-c("Total","Interpretation required","No")
```

> m

	Total	Interpretation	required	No
Total	9189		381	8808
Kinh	7745		0	7745
Non-kinh	1444		381	1063

6.2 Religion

Ethnicity and religion are the very important factors when analyzing social and cultural structure of the country. In VHLSS, question on religion is not found in the questionnaire.

According to 2009 Population Census, only a small fraction of the Vietnamese adheres to institutional religions;

Non religious,	80.8%
Buddhism,	9.3%
Christianity,	7.2%
Roman Catholic,	6.7%
Protestant,	0.5%
Hòa Hảo,	1.5%
Cao Đài,	1.1%

"Although according to a 1999 census most Vietnamese list themselves as having no religious affiliation, religion, as defined by shared beliefs and practices, remains an integral part of Vietnamese life, dictating the social behaviours and spiritual practices of Vietnamese individuals in Vietnam and abroad. The triple religion (Vietnamese: tam giáo), referring to the syncretic combination of Mahayana Buddhism, Confucianism, and Taoism remains a strong influence on the beliefs and practices of the Vietnamese, even if the levels of formal membership in these religious communities may not reflect that influence. One of the most notable and universal spiritual practices common to Vietnamese is ancestor veneration, a practice shared with Chinese and most other Asian cultures. Practically all Vietnamese, regardless of formal religious affiliation, have an altar in their home or business where prayers are offered to their ancestors. These offerings and practices are done frequently during important traditional or religious celebrations (e.g., death anniversaries), the starting of a new business, or even when a family member needs guidance or counsel. Belief in ghosts and spirits is common; it is commonly believed that failing to perform the proper rituals for one's ancestors will cause them to become hungry ghosts (Vietnamese: ma đói)." (Wikipedia)

6.3 Age heaping

Age heaping is popular in developing countries' censuses and surveys.

The below chart shows the number of sample household members of VHLSS 2006 by sex and single year of age.

It is understood that age heaping is not significant in Vietnam. The reason might be that they use twelve signs of Chinese Zodiac in daily life, which is listed in the questionnaire.



Chart Number of sample household members by sex and single year of age (2006)

> t<-tapply(muc1a\$PID,list(muc1a\$m1ac5,muc1a\$m1ac2),length)

> dim(t)

[1] 104 2

> colnames(t)<-c("Male","Female")</pre>

> m<-data.frame(age=rownames(t),t)

> head(m)

age Male Female

- 0 0 224 229
- 1 1 265 224
- 2 2 254 220
- 3 3 239 267
- 4 4 238 210
- 5 5 273 253

> write.csv(m,"pop_by_age_sex.csv",row.names=F)

TABLE OF THE LUNAR-SOLAR CALENDAR YEAR EQUIVALENCE

Mouse	1900	1912	1924	1936	1948	1960	1972	1984	1996
Buffalo	1901	1913	1925	1937	1949	1961	1973	1985	1997
Tiger	1902	1914	1926	193B	1950	1962	1974	1986	1998
Cat	1903	1915	1927	1939	1951	1963	1975	1987	1999
Dragon	1904	1916	1928	1940	1952	1964	1976	1988	2000
Snake	1905	1917	1929	1941	1953	1965	1977	1989	2001
Horse	1906	1918	1930	1942	1954	1966	1978	1990	2002
Goat	1907	1919	1931	1943	1955	1967	1979	1991	2003
Monkey	1908	1920	1932	1944	1956	1968	1980	1992	2004
Cock	1909	1921	1933	1945	1957	1969	1981	1993	2005
Dog	1910	1922	1934	1946	1958	1970	1982	1994	2006
Pig	1911	1923	1935	1947	1959	1971	1983	1995	

Note: Buffalo is the second symbol, taking place of Ox in the Chinese zodiac. Also the fourth symbol of Rabbit is replaced with Cat.

6.4 Literacy

✓ In section 2A (Education) of the questionnaire, literacy was asked for those who finished 4-th grade or below, and never go to school. Those who finished 5-th grade (primary school) or upward are regarded as literacy.

 \checkmark Generated the variable literate for all household members

1: literate

2: Illiterate

> d<-muc2a

```
> literate<-ifelse(dm2ac1>=5|dm2ac2==1, 1, 2)
```

```
# Number of records by literacy
> table(literate)
literate
    1    2
33211   5860
```

✓ Literacy rate of Vietnam is 86%.

```
> round (prop. table(tapply(d$wt, literate, sum, na. rm=T))*100, 1)
```

1 2

86.1 13.9

✓ Merged literate with MUC1A, and appended data of urban/rural and ethnicity in TTCHUNG.

> df<-merge(cbind(muc1a, literate), ttchung[, c("ID", "ttnt", "dantoc")], key="ID", all.x=T)
> df["kinh"]<-ifelse(df\$dantoc==1, 1, 2)</pre>
✓ Literacy rate by region and urban/rural

Literacy rate is low in rural of North West (67%) and rural of Central Highlands (77%).

> denominator<-addmargins(tapply(df\$wt, list(substr(df\$tinh, 1, 1), df\$ttnt), sum, na.rm=T))

> d2<-subset(df, literate==1)</pre>

- > numerator<-addmargins(tapply(d2\$wt,list(substr(d2\$tinh,1,1),d2\$ttnt),sum,na.rm=T))
- > rate<-round (numerator/denominator*100, 1) [c (9, 1:8), c (3, 1, 2)]
- > colnames(rate)<-c("Total", "Urban", "Rural")
- > rownames(rate) <-c("Vietnam", region.name)
- > rate

	Total	Urban	Rural
Vietnam	86. 1	90.0	84. 7
Red River Delta	90. 0	91.9	89.4
North East	85. 8	92.3	84. 2
North West	70. 5	90. 7	67. 2
North Central	87. 2	92.6	86.3
South Central Coast	87.4	89.5	86.5
Central Highlands	79. 7	87.3	76. 5
South East	88. 2	90.6	85.7
Mekong River Delta	83. 6	85.8	83. 1

✓ Literacy rate by ethnicity

Literacy rate of non-Kinh people is 15% points lower than Kinh people. Range of Kinh people's literacy rates by region is narrower compared to non-Kinh people.

- > denominator<-addmargins(tapply(df\$wt, list(substr(df\$tinh, 1, 1), df\$kinh), sum, na.rm=T))
- > d2<-subset(df, literate==1)</pre>
- > numerator<-addmargins(tapply(d2\$wt,list(substr(d2\$tinh,1,1),d2\$kinh),sum,na.rm=T))
- > rate<-round (numerator/denominator*100, 1) [c (9, 1:8), c (3, 1, 2)]
- > colnames(rate)<-c("Total", "Kinh", "Non-kinh")
- > rownames(rate)<-c("Vietnam", region.name)

> rate

	Total	Kinh	Non-kinh
Vietnam	86. 1	88.5	72. 3
Red River Delta	90. 0	90.0	88. 9

North East	85.8 90.3	79.8
North West	70.5 93.3	65.3
North Central	87. 2 88. 4	76. 7
South Central Coast	87.4 88.9	64. 5
Central Highlands	79.7 90.4	59. 2
South East	88. 2 89. 2	77.9
Mekong River Delta	83.6 84.8	69. 2

6.5 Polygamy

Summary of Marriage System

- VHLSS allows the cases of multiple spouses. Examples of micro data of VHLSS 2006 are shown later.
- The enumerator's manual "Operational Handbook of VHLSS 2004" describes household members as follows;

"If the household owner has many wives, then enumerators write the names of the first wife and her children, then the second wife and her children and go on in this way." (at page 21)

- In Vietnam, men belonging to the old generation (born around 1930s or earlier) usually had more than one wife. That was the customs of those days.
- According to the Law of Marriage and Family, which has begun since 1959 in North Vietnam at the time and was applied for South Vietnam after the end of War in 1975, one husband has one wife only. Until now, this provision is still in effect. Therefore, people got married before the implementation of the Law could have more spouses if their marriage was carried out before that time.
- Vietnam has the Law of Family Register System. Accordingly, the second spouse cannot be registered as an official spouse at the Local Authority. However, VHLSS is flexible and still accepts the second spouse if the family answer that they are spouses.

VHLSS 2006 – Example of polygamy

- ✓ Data file MUC1A: List of household members
- > head (muc1a)
- tinh huyen xa diaban hoso matv m1ac2 m1ac3 m1ac4a m1ac4b m1ac5 m1ac6 m1ac7

1	101	01 03	014	15	1	2	1	7	1957	48	3	12
2	101	01 03	014	15	2	1	3	11	1977	28	1	12
3	101	01 03	014	15	3	1	3	3	1980	26	1	12
4	101	01 03	014	15	4	1	3	3	1986	20	1	12
5	101	01 03	014	19	1	2	1	1	1952	54	3	12

6	101	01 C	03 014	4 19	2	1		3	1	1980	26	2		12
	m1ac8	m1ac9	m1ac10a	m1ac10b		I	D		wt	xaid			PIC)
1	1	NA	NA	NA	101	01030141	5	3107.	318	1010103	101010	03014	1501	
2	1	NA	NA	NA	101	01030141	5	3107.	318	1010103	101010	03014	1502	2
3	1	NA	NA	NA	101	01030141	5	3107.	318	1010103	101010	03014	1503	3
4	1	NA	NA	NA	101	01030141	5	3107.	318	1010103	101010	03014	1504	1
5	1	NA	NA	NA	101	01030141	9	3107.	318	1010103	101010	03014	1901	
6	1	NA	NA	NA	101	01030141	9	3107.	318	1010103	101010	03014	1902	2

- ✓ Generated subset of spouse.
- > df <-muc1a[, c(18, 6:8, 11, 12)]

```
> colnames(df)[3:6]<-c("sex", "relation", "age", "marital")</pre>
```

> head (df)

ID matv sex relation age marital

1	101010301415	1	2	1	48	3
2	101010301415	2	1	3	28	1
3	101010301415	3	1	3	26	1
4	101010301415	4	1	3	20	1
5	101010301419	1	2	1	54	3
6	101010301419	2	1	3	26	2

```
> spouse<-subset(df,relation==2)
```

 \checkmark Made frequency table of the number of spouse within the household.

The results show that there are 4 sample households with two spouses.

```
> ns<-(tapply(spouse$ID, spouse$ID, length))
```

```
> table(ns)
ns
1 2
7340 4
```

 \checkmark Example of household with multiple spouses within the household

```
# IDs of such households
```

> names (ns) [ns==2]

[1] "103117100913" "105253900215" "109060100814" "201150700215"

```
> df[df$ID=="103117100913",]
```

	ID	matv	sex	relation	age	marital
1333	103117100913	1	1	1	80	2
1334	103117100913	2	2	2	78	2
1335	103117100913	3	2	2	63	2
1336	103117100913	4	1	3	27	2
1337	103117100913	5	2	3	25	2
1338	103117100913	6	1	6	6	NA
1339	103117100913	7	1	6	3	NA
> df	[df\$ID=="10525	539002	215″,]		
	ID	matv	sex	relation	age	marital
3137	105253900215	1	1	1	42	2
3138	105253900215	2	2	2	34	2
3139	105253900215	3	2	3	11	NA
3140	105253900215	4	2	2	30	2
3141	105253900215	5	1	3	5	NA
3142	105253900215	6	2	3	2	NA
> df	[df\$ID=="10906	601008	314″,]		
	ID	matv	sex	relation	age	marital
4692	109060100814	1	1	1	47	2
4693	109060100814	2	2	2	49	2
4694	109060100814	3	2	2	46	2
4695	109060100814	4	1	3	14	1
4696	109060100814	5	2	3	12	NA
> df	[df\$ID=="20118	507002	215″,]		
	ID	matv	sex	relation	age	marital
7705	201150700215	1	1	1	33	2
7706	201150700215	2	2	2	42	2
7707	201150700215	3	2	2	43	2
7708	201150700215	4	1	3	18	2
7709	201150700215	5	2	3	11	NA
7710	201150700215	6	2	3	7	NA
7711	201150700215	7	2	3	15	2
7712	201150700215	8	1	7	13	1
7713	201150700215	9	2	7	8	NA
7714	201150700215	10	1	7	9	NA
7715	201150700215	11	2	3	4	NA

7716 201150700215 12 1 3 2 NA

Among 4 cases, the first case with spouses' ages 63 years old and 78 years. Remaining 3 cases, location of which are Red River Delta and North East regions, are spouses with ages around 40-50 years old.

According to Ms. Van,

"Therefore, the first case may belong to the custom of old generation. The other 3 cases may be they violate the Law, but with some special reasons (for example, to have children if the first wife could not produce kids, ...). Because of the preference of son, if the wife has the daughters only, the man could consider to have the second wife, though it is not permitted by the Law."

6.6 Relationship to the household head

Summary

• According to "Operational Handbook of VHLSS 2004", the variable of person number MATV

follows the next rule;

- The first person to be written down in the questionnaire is the household owner although this person may not be the respondent or even when this person is not present at home. Household owner is always coded number 1.

- Then followed by household owner's spouse and children who are not yet married from the oldest one to the youngest. If the household owner has many wives, then enumerators write the names of the first wife and her children, then the second wife and her children and go on in this way.

- Then followed by household owner's children who are married, their spouse and children (if any).

- Then come father, mother, adopted brothers and sisters, maternal and paternal grandparents, maternal and paternal grandchildren (whose parents are both not living in the surveyed household) and other relatives of the household owner and his/her spouse.

- Finally, non-relatives.

• The code of relationship differs between VHLSS 2002 and 2004/2006.

For VHLSS 2002

Code	Relationship to head
1	Head
2	Wife/husband
3	Child
4	Child in law
5	Parents
6	Sister/brother
7	Grandfather/grandmother
8	Grandchild
9	Other relationship

For VHLSS 2004 and 2006

Code	Relationship to head
1	Head
2	Spouse
3	Children

4	Parents
5	Grandfather/Grandmother
6	Grandchild
7	Other relation
9	Missing

The code for "Child in law", that is, son's wife or daughter's husband, seems not to be clearly described in the manual of VHLSS 2004.

For this, Ms. Van describes, "Explaining for enumerator is that code 3 is the children of the head. In other words, code 3 is for the one who is blood/adopted children of the head. Code 7 is for other relationship."

However, micro data of VHLSS 2006 suggests that there are two cases of code=3 (child) and code=7 (other relation). Next are examples.

```
> df[df$ID=="715032100614",]
               ID matv sex relation age marital
30336 715032100614
                   1
                       1
                                 1 50
                                            2
30337 715032100614
                    2
                        2
                                 2 53
                                            2
30338 715032100614
                    3 1
                                 3 18
                                            1
30339 715032100614
                   4 1
                                 3 15
                                            1
30340 715032100614
                                 3 14
                    52
                                            1
30341 715032100614
                                 3 29
                                            2
                    6
                       1
30342 715032100614
                    7
                        2
                                 3 26
                                            2
30343 715032100614
                    8 1
                                 6
                                    7
                                            NA
30344 715032100614
                                 6
                                    5
                    9
                        1
                                            NA
30345 715032100614
                   10
                                 6
                                    2
                                            NA
                       1
30346 715032100614
                   11
                        1
                                 6
                                    1
                                            NA
30347 715032100614
                        2
                                 3 26
                                            2
                   12
30348 715032100614
                                 3
                                    28
                                            2
                   13
                        1
30349 715032100614
                   14
                        1
                                 6
                                    3
                                            NA
30350 715032100614
                   15
                        1
                                 3
                                    21
                                            2
30351 715032100614
                                            2
                   16
                        2
                                 3 21
30352 715032100614
                   17
                        2
                                 4 76
                                            3
```

In the above case of relationship code=3, MATV=7 and 16 are regarded as sons' wives, and MATV=13 is a daughter's husband.

> df[df\$ID==~303090700214",] ID matv sex relation age marital 14437 303090700214 1 1 1 56 2

14438	303090700214	2	2	2	53	2
14439	303090700214	3	1	3	16	1
14440	303090700214	4	2	3	14	1
14441	303090700214	5	1	3	22	2
14442	303090700214	6	2	7	23	2
14443	303090700214	7	2	6	6	NA
14444	303090700214	8	2	6	4	NA
14445	303090700214	9	2	6	1	NA
14446	303090700214	10	1	3	20	2
14447	303090700214	11	2	7	22	2
14448	303090700214	12	1	6	4	NA
14449	303090700214	13	1	6	2	NA
14450	303090700214	14	1	6	0	NA
14451	303090700214	15	2	7	22	2
14452	303090700214	16	1	3	18	2
14453	303090700214	17	2	7	20	2

In the above case of relationship code=7, MATV=6, 11 and 17 are regarded as sons' wives. MATV=15 is supposed to be a son's wife while the son (her husband) is away from home.

In summary,

Types of code for spouse of son/daughter	Number of sample households
Case 1: code 3 was applied for daughter's husband.	115
Case 1: code 3 was applied for son's wife.	668
Case 2: code 7 was applied for daughter's husband.	36
Case 2: code 7 was applied for son's wife.	289

- # Made use of data frame ft, which was created in 6.7.1.
- > child.couple<-function(df,k,j) {
- # return 2 if the code of son/daughter' s spouse is 3.
- # return 4 if the code of son/daughter' s spouse is 4.
- + flag<-0
- + if(j>=2&df[k, j+27]==2&df[k, j+28]==2&df[k, j+44]*df[k, j+45]==2&
- + (df[k, j+10]==3&df[k, j+11]==3)) {
- + flag<-2
- + }else if(j>=2&df[k, j+27]==2&df[k, j+28]==2&df[k, j+44]*df[k, j+45]==2&
- + (df[k, j+10] == 3&df[k, j+11] == 7)) {
- + flag<-4
- + }
- + return(flag)

+ }

```
> childinlaw<-data.frame(ID=ft$ID,matrix(0,nrow=nrow(ft),ncol=5))
> colnames (childinlaw) <- c ("ID", "m3", "f3", "m7", "f7", "other")
> for(k in 1:nrow(ft)) {
+ for(j in 1:ft[k, 2]){
# Add 1 to childinlaw() if the spouse if female
+ x<-child.couple(ft,k,j)+(ft[k,j+45]==2)
+ if(x>0) childinlaw[k,x]<-childinlaw[k,x]+1
+ }
+ }
> sum(childinlaw$m3)
[1] 115
> sum(childinlaw$f3)
[1] 668
> sum(childinlaw$m7)
[1] 36
> sum(childinlaw$f7)
[1] 289
> sum(childinlaw$other)
[1] 0
```

- When discussing the family type, the sequence of person number MATV is indispensable for determining a couple.
- The two household members within the household who satisfy the next conditions are presumed as a couple in VHLSS 2006.
 - 1) Their person numbers (MATV) are consecutive.
 - 2) Their marital statuses are married.
 - 3) Their sexes are opposite.
 - Their relationship codes are the same, or the first person's code is 3 (Child) and the second person's code is 7 (Other relation).
- "Brother/sister", "other relatives" and "non-relatives" are not distinguished clearly in the questionnaire.

Background of the relationship codes

It is said that the traditional Vietnamese family is patriarchal, especially among Kinh people. Traditional values of Vietnamese lifestyle were deeply affected by Confucian ethics. During thousand years the Chinese invaded and maintained control Vietnam, Vietnamese culture was permeated by their Confucian philosophical beliefs.

Regarding the reason why the response categories of relationship was simplified in VHLSS 2004/2006, Ms. Van replied, "We distinguish the direct relationship within family. The direct relationship with the head is important to be distinguished clearly."

For reference:

Relationship to the household head in Population Census

- In Vietnam 1999 Population Census, the response categories of the question on the relationship to the household head are as follows. <u>Code 3 "Son/daughter" is limited only for biological</u> <u>child.</u>
 - 1 Household head
 - 2 Spouse
 - 3 Son/Daughter
 - 4 Parents
 - 5 Other
- The definition of each response category is described in the Enumerators' manual;

Question 2: What is the relationship of (name) to the household head?

If respondent is the head of household, interviewers fill in code "1". For the rest of household members, interviewers fill in corresponding codes 2 to 5 depending on their relationship with the head of household. Definition of each relationship to head of household is, as follow:

<u>Head of household</u>: is a household representative who is recognized by all members of household. In households where there are only children because their father and mother work in ministry of defense and/or ministry of public security (enumerating in special plan), head of household is the oldest child. For students of professional schools or college, students of stay-in general schools, interviewers register them by their room/apartment as one household. Head of household is a person who is recognized by all household's members. Other household's members would have "other" relationship to the head. <u>Spouse</u>: if a person has two spouses or more (wives or husbands) and lives together with them in one household, these persons are all considered as spouses (wives or husbands) of the head of household (see question 15). Interviewers select code "2" if a person is a spouse of the head of household. <u>Biological child</u>: is child by birth of the head of household. Interviewers select code "3" if a person is defined as biological child.

<u>Parent</u>: Father (mother) of the head of household includes biological father (mother), adopted father (mother), father-in-law and mother-in-law of the head of household.

<u>Other relationship</u>: Interviewers should report specific relationship with the head of household, such as stepchild of the head of household, daughter-in-law/son-in-law, adopted child, other brother, older sister, younger brother/sister, aunt, uncle, grandparent, grandchild, domestic employees, maid, friend, etc. A person is defined as "other relationship" with the head of household has code "5".

6.7 Family type

It is said that the Vietnamese household traditionally followed the extended multi-generational pattern. The parents, their sons and their wives, their children, and unmarried siblings usually constituted a Vietnamese household.

Objectives:

- 1) To determine the family type of each household.
- 2) To analyze living standards by family type.

6.7.1 Family type based on the number of persons by generation

Data file for determining the family type (1)

✓ Designed household-level data file consist of number of persons by relationship as well as data of relationship, marital status, sex and age of each person.

Layout of record

ID	Number of persons by relationship							Re	latio	nshij	o of e	ach	Ма	rital	st	atus	of	Se	x of	each	n pers	on	Ag	e of	each	pers	on			
								person					each person																	
	Т	1	2	3	4	5	6	7	х	9	1	2	3		17	1	2	3		17	1	2	3		17	1	2	3		17

✓ Verified MUC1A

- > df<-muc1a[, c(18, 6:8, 11, 12)]
- > colnames(df)[3:6]<-c("sex", "relation", "age", "marital")
- > table(df\$sex,useNA="ifany")
 - 1 2
- 19157 19914
- > table(df\$relation, useNA="ifany")
- 1 2 3 4 5 6 7 9
- 9189 7348 18275 952 45 2203 1057 2
- > table(subset(df, age>=13)\$marital, useNA="ifany")

1 2 3 4 5 10880 18029 2028 265 135 > table(is.na(df\$age)) FALSE 39071

✓ Created data frame ft: family type

household size

- > hhsize<-tapply(df\$ID, df\$ID, length)
- > ft<-data.frame(ID=names(hhsize), hhsz=hhsize, row.names=NULL)
- # Number of persons by relationship
- > for(j in c(1:7,9)) {
- + d<-subset(df,relation==j)
- + nj<-tapply(d\$ID, d\$ID, length)
- + dj<-data.frame(ID=names(nj),nj)
- + ft<-merge(ft,dj,by.x="ID",by.y="ID",a||.x=T)
- + colnames(ft)[ncol(ft)]<-paste("n",j,sep="")
- + }
- > dim(ft)
- [1] 9189 10

```
> head(ft)
```

ID hhsz n1 n2 n3 n4 n5 n6 n7 n9

1	101010301415	4	1	NA	3	NA	NA	NA	NA	NA	
2	101010301419	4	1	NA	3	NA	NA	NA	NA	NA	
3	101010301424	4	1	NA	NA	NA	NA	NA	3	NA	
4	101010901913	2	1	1	NA	NA	NA	NA	NA	NA	
5	101010901915	3	1	1	1	NA	NA	NA	NA	NA	
6	101010901919	3	1	1	1	NA	NA	NA	NA	NA	

> ft.old<-ft

Relationship, marital status, sex and age by person number

- > for(j in 1:17){
- + d<-subset(df,matv==j)[,c(1,4,6,3,5)]
- + colnames(d) <-c("ID", paste(c("r", "m", "s", "a"), j, sep=""))

Vietnam VHLSS Manual - Overall and survey process (Version 1.1)

[61] "s13" ″r15″ ″s15″ ″a13″ ″r14″ ‴m14″ ″s14″ ″a14″ ‴m15″ ″a15″ ‴a17″ ‴m16″ ″s16″ ″a16″ ″r17″ ‴m17″ ″s17″ [71] "r16" # Ordered the variables as designed > ft<-ft[, c(1:10, seq(11, 75, by=4), seq(12, 76, by=4), seq(13, 77, by=4), seq(14, 78, by=4))] > colnames(ft) [1] ″ID″ "hhsz" "n1" ″n3″ ″n4″ ″n7″ ″n2″ ″n5″ ″n6″ ″n9″ [11] "r1" ″r2″ ″r3″ ″r4″ ″r5″ ″r6″ ″r7″ ″r8″ ″r9″ ″r10″ ‴r13″ ″r16″ [21] "r11" ″r12″ ″r14″ ″r15″ ″r17″ ″m1″ ‴m2″ ″m3″ [31] "m4" ″m5″ ″m6″ ″m7″ ″m8″ ‴m9″ ‴m10″ ″m11″ ‴m12″ ‴m13″ ″s4″ ″s5″ ″s6″ [41] "m14" ″m15″ ″m16″ ‴m17″ ″s1″ ″s2″ ″s3″ [51] ″s7″ ″s8″ ″s9″ ″s10″ ″s11″ ″s12″ ″s13″ ″s14″ ″s15″ ″s16″ [61] ″s17″ ″a1″ ‴a2″ ″a3″ ″a4″ ″a5″ ″a6″ ″a7″ ″a8″ ″a9″ [71] "a10" "a11" "a12" "a13" "a14" ″a15″ ″a16″ ″a17″ > head(ft[, c(1:16, 28:33, 45:50, 62:67)]) ID hhsz n1 n2 n3 n4 n5 n6 n7 n9 r1 r2 r3 r4 r5 r6 m1 m2 m3 m4 m5 m6 s1 s2 s3 1 101010301415 4 1 NA 3 NA NA NA NA NA 1 3 3 3 NA NA 3 1 1 1 NA NA 2 1 1 2 101010301419 4 1 NA 3 NA NA NA NA NA 1 3 3 3 NA NA 3 2 2 1 NA NA 2 1 2 3 101010301424 4 1 NA NA NA NA NA 3 NA 1 7 7 7 NA NA 3 2 2 NA NA NA 2 1 2 4 101010901913 2 1 1 NA NA NA NA NA NA 1 2 NA NA NA NA 2 2 NA NA NA 1 2 NA 5 101010901915 3 1 1 1 NA NA NA NA NA 1 2 3 NA NA NA 2 2 1 NA NA NA 1 2 2 6 101010901919 3 1 1 1 NA NA NA NA NA 1 2 3 NA NA NA 2 2 1 NA NA NA 2 1 2 s4 s5 s6 a1 a2 a3 a4 a5 a6 1 1 NA NA 48 28 26 20 NA NA

- > colnames(ft) [1] "ID" "hhsz" "n1" ″n2″ ″n3″ ″n4″ ″n5″ ″n6″ ″n7″ ″n9″ [11] "r1" ″m1″ ″s1″ ″a1″ ″r2″ ‴m2″ ″s2″ ″a2″ ″r3″ ″m3″ [21] "s3" ″a3″ ″r4″ ‴m4″ ″s4″ ″a4″ ″r5″ ″s5″ ″m5″ ″a5″ ″a7″ [31] ″r6″ ″m6″ ″s6″ ″a6″ ″r7″ ″m7″ ″s7″ ″r8″ ″m8″ ‴m9″ ″a9″ [41] ″s8″ ″a8″ ″r9″ ″s9″ ″r10″ ″m10″ ″s10″ ″a10″ [51] "r11" ″m11″ ″s11″ ″a11″ ″r12″ ″s12″ ″a12″ ″r13″ ‴m13″ ‴m12″
- > ft.old2<-ft
- [1] 9189 78
- > dim(ft)
- + }
- + ft<-merge(ft, d, by. x="ID", by. y="ID", all. x=T)

87

2 1 NA NA 54 26 25 19 NA NA

3 1 NA NA 74 28 24 1 NA NA

4 NA NA NA 77 74 NA NA NA NA

5 NA NA NA 48 39 20 NA NA NA

6 NA NA NA 62 59 21 NA NA NA

> ft.old3<-ft

- > ft[is.na(ft)]<-0
- > head(ft[, c(1:16, 28:33, 45:50, 62:67)])

				II) hl	hsz	n1	n2	n3	n4	n5	n6	n7	n9	r1	r2	r3	r4	r5	r6	m1	m2	m3	m4	m5	m6	s1	s2	s3
1	101	1010)30 [.]	141	5	4	1	0	3	0	0	0	0	0	1	3	3	3	0	0	3	1	1	1	0	0	2	1	1
2	101	1010)30 [.]	1419	9	4	1	0	3	0	0	0	0	0	1	3	3	3	0	0	3	2	2	1	0	0	2	1	2
3	101	1010	030	1424	4	4	1	0	0	0	0	0	3	0	1	7	7	7	0	0	3	2	2	0	0	0	2	1	2
4	101	101(090 ⁻	191:	3	2	1	1	0	0	0	0	0	0	1	2	0	0	0	0	2	2	0	0	0	0	1	2	0
5	101	101(090 ⁻	191	5	3	1	1	1	0	0	0	0	0	1	2	3	0	0	0	2	2	1	0	0	0	1	2	2
6	101	1010	090 [.]	1919	9	3	1	1	1	0	0	0	0	0	1	2	3	0	0	0	2	2	1	0	0	0	2	1	2
	s4	s5	s6	a1	a2	a3	a4	a5	a6																				
1	1	0	0	48	28	26	20	0	0																				
2	1	0	0	54	26	25	19	0	0																				
3	1	0	0	74	28	24	1	0	0																				
4	0	0	0	77	74	0	0	0	0																				
5	0	0	0	48	39	20	0	0	0																				
6	0	0	0	62	59	21	0	0	0																				

Data file for determining the family type (2)

Designed household-level data file consist of number of persons by generation as well as \checkmark number of couples by generation.

For each household;

Genera	ation based on the household head	Number of persons	Number of couples
1	Grandparents	p1	c1
2	Parents	p2	c2
3	Head/spouse	p3	c3

4	Children	p4	c4
5	Grandchildren	p5	c5
6	Brother/sister	рб	сб
7	Other	p7	c7

✓ Defined function couple(df, k, j)

df: data frame of family type, defined in the above

k: record number of df

j: person number of df

- couple=1 if j-th and (j+1)-th persons are a couple
- couple=0 else
- The two household members within the household who satisfy the next conditions are presumed as a couple in VHLSS 2006.
 - 1) Their person numbers (MATV) are consecutive.
 - 2) Their marital statuses are married.
 - 3) Their sexes are opposite.
 - 4) Their relationship codes are the same, or

the first person's code is 3 (Child) and the second person's code is 7 (Other relation).

```
> couple<-function(df,k,j) {
```

+ flag<-0

```
+ if(j==1&df[k, j+11]==2) {
```

- + f∣ag<-1
- + return(flag)
- + } else if(j>=2&df[k, j+27]==2&df[k, j+28]==2&df[k, j+44]*df[k, j+45]==2&
- + (df[k, j+10] == df[k, j+11] | (df[k, j+10] == 3&df[k, j+11] == 7)))
- + flag<-1
- + }
- + return(flag)

```
+ }
```

```
Examples;
```

```
> couple(ft, 2, 2)
[1] 1
```

```
> couple(ft, 5, 1)
[1] 1
> couple(ft, 1, 1)
[1] 0
> couple(ft, 3, 1)
[1] 0
> couple(ft, 3, 2)
[1] 1
```

✓ Defined function generation(df, k, j)

df: data frame of family type, defined in the above

k: record number of df

j: person number of df

generation: j-th person's generation code defined in the above table

> generation<-function(df, k, j) {

+ g<−7

```
+ if(df[k, j+10]==1|df[k, j+10]==2) g<-3
```

```
+ else if(df[k,j+10]==3) g<-4
```

```
+ else if(df[k,j+10]==4) g<-2
```

```
+ else if(df[k, j+10]==5) g<-1
```

- + else if(df[k, j+10]==6) g<-5
- + # Child in law
- + else if (df[k, j+10]==7&df[k, j+9]==3&couple(df, k, j-1)==1) g<-4
- + # Brother/sister
- + flag<-0
- + for (n in (j+1):df[k, 2]) {
- + if(df[k, n+10]==5) flag<-flag+1
- + } # end of for
- + if(g==7&f|ag>0) g<-6
- + return(g)

```
+ } # end of function
```

Remarks:

Code=6 (Brother/sister) is identified only if Grandparents exist in the household.

A part of brothers/sisters might be grouped as code=7.

✓ Created household-level data file ft2 consist of number of persons by generation as well as number of couples by generation.

```
> ft2<-data.frame(ID=ft$ID.matrix(0.nrow=nrow(ft).ncol=14))
> colnames(ft2)<-c("ID", paste("p", 1:7, sep=""), paste("c", 1:7, sep=""))
> for (k in 1:nrow(ft)) {
+ for(j in 1:ft[k, 2]) {
+ g<-generation(ft, k, j)
+ ft2[k,g+1]<-ft2[k,g+1]+1
+ ft2[k,g+8]<-ft2[k,g+8]+couple(ft,k,j)
+ }
+ }
> head(ft2)
              ID p1 p2 p3 p4 p5 p6 p7 c1 c2 c3 c4 c5 c6 c7
1 101010301415 0 0 1 3 0 0 0 0 0 0 0 0 0 0
2 101010301419 0 0 1 3 0 0 0 0 0 0 1 0 0 0
3 \hspace{.15cm} 101010301424 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 1 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 3 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 0 \hspace{.15cm} 1 \hspace{.15cm} 1
4 101010901913 0 0 2 0 0 0 0 0 0 1 0 0 0 0
5 101010901915 0 0 2 1 0 0 0 0 0 1 0 0 0 0
6 101010901919 0 0 2 1 0 0 0 0 0 1 0 0 0 0
> ft2.old<-ft2
# Generated the variable of hhsz: household size
> ft2["hhsz"]<-rowSums(ft2[, 2:8])</pre>
# Number of persons by generation
> sapply(ft2[, 2:8], sum)
   p1
          p2
                 p3
                                      p6
                             p5
                                             p7
                        p4
   45
         952 16537 18600 2203
                                        4
                                            730
# Generated the variable of ct: total number of married couples with household
> ft2["ct"]<-rowSums(ft2[, 9:15])</pre>
```

Number of married couples by generation

> sapply(ft2[,9:15],sum)
c1 c2 c3 c4 c5 c6 c7
2 115 7331 1109 3 0 40

 $\#\ {\rm Frequency}\ {\rm of}\ {\rm sample}\ {\rm households}\ {\rm by}\ {\rm total}\ {\rm number}\ {\rm of}\ {\rm married}\ {\rm couples}\ {\rm within}\ {\rm household}$

> addmargins(table(ft2\$ct))

0 1 2 3 4 5 Sum 1384 7068 687 43 6 1 9189

Grouping sample household based on the number of persons by generation

✓ Excluded 482 households with p7>0, and grouped 8,707 rest of households based on the number of household members by generation.

Two-generation household is the majority, which occupies 65%, followed by three-generation household (19%) and one-generation household (10%).

Number of	Household consist of the below generations	Code	Number of	Percentage
generation			households	(%)
	All households		9,189	100.0
1	Total	h10	914	9.9
	Single	h11	344	3.7
	Plural members (head, spouse only)	h12	570	6.2
	Plural members including siblings	h13	0	-
2	Total	h20	5,927	64.5
	Head - Children	h21	5,855	63.7
	Parents - Head	h22	72	0.8
3	Total	h30	1,779	19.4
	Head – Children – Grandchildren	h31	1,074	11.7
	Head - Grandchildren	h32	96	1.0
	Parents – Head – Children	h33	606	6.6
	Grandparents – Parents - Head	h34	0	-
	Grandparents - Head	h35	3	0.0
4	Total	h40	84	0.9

Table Type of households by number of household members' generations

	Parents – Head – Children - Grandchildren	h41	46	0.5
	Parents – Head –Grandchildren	h42	4	0.0
	Grandparents – Parents – Head - Children	h43	3	0.0
	Grandparents –Head - Children	h44	31	0.3
5	Grandparents – (Parents) – Head – (Children) – Grandchildren	h50	3	0.0
x	Households with other relatives and/or non-relatives	h60	482	5.2

Generated the variable code of household type in data frame ft2.

> d<-subset(ft2, p7==0)

> d["code"]<-0

> d["code"]<-ifelse(d\$p3==1& d\$hhsz==1, 11, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2==0& d\$p3>=2& d\$p4==0& d\$p5==0& d\$p6==0, 12, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2==0& d\$p3>0& d\$p4==0& d\$p5==0&d\$p6>0, 13, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2==0& d\$p4>0& d\$p5==0, 21, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2>0& d\$p4==0& d\$p5==0, 22, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2==0& d\$p4>0& d\$p5>0, 31, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2==0& d\$p4==0& d\$p5>0, 32, d\$code)

 $\label{eq:code} > \ d["code"] <\!\!-ifelse(dp1==0\&\ dp2>0\&\ dp4>0\&\ dp5==0,\ 33,\ dcode) \\$

> d["code"] <-ifelse(d\$p1>0& d\$p2>0& d\$p4==0& d\$p5==0, 34, d\$code)

 $\label{eq:code} > \ d["code"] <- ifelse (dp1>0& dp2==0& dp3==0& dp5==0, 35, dcode) \\$

> d["code"]<-ifelse(d\$p1==0& d\$p2>0& d\$p4>0& d\$p5>0, 41, d\$code)

> d["code"]<-ifelse(d\$p1==0& d\$p2>0& d\$p4==0& d\$p5>0, 42, d\$code)

> d["code"]<-ifelse(d\$p1>0& d\$p2>0& d\$p4>0& d\$p5==0, 43, d\$code)

> d["code"]<-ifelse(d\$p1>0& d\$p2==0& d\$p4>0& d\$p5==0, 44, d\$code)

> d["code"]<-ifelse(d\$p1>0& d\$p5>0, 50, d\$code)

> (t<-table(d\$code))

11	12	21	22	31	32	33	35	41	42	43	44	50
344	570 క	5855	72 1	074	96	606	3	46	4	3	31	3

> sum(t)

[1] 8707

> round (t/nrow(ft2)*100, 1)

11	12	21	22	31	32	33	35	41	42	43	44	50
3.7	6.26	3.7	0.81	1.7	1.0	6.6	0.0	0.5	0.0	0.0	0.3	0.0

Added code to ft2 > ft2<-merge(ft2,d[c("ID", "code")], by="ID", all=T) > ft2[is.na(ft2\$code),"code"]<-60 > table(ft2\$code,useNA="ifany") 33 11 12 21 22 31 32 35 41 42 43 44 50 60 344 570 5855 72 1074 96 606 3 46 4 3 31 3 482 > ft2.old2<-ft2

6.7.2 Family type based on the generation and number of married couples

Grouping sample household based on the generation and number of married couples

✓ Excluded 482 sample households with other relation and 1,262 households without married couple, then grouped 7,445 rest of households by the number of married couples' generation and number of couples.

Nuclear family household is the major type, which occupies 61%, followed by the below types;

- ➢ Household consist of head's couple and children's couple (6.5%)
- ➢ Household consist of head's couple, their parent(s) and their child(ren) (5.3%)
- \blacktriangleright Household without head's couple but with children's couple(s) (4.4%)
- ➢ Household consist of head's couple, their child(ren) and their grandchild(ren) (1.9%)

As for number of married couples' generation, households including one-generation couple(s) is 74%, and household including two-generation couples is 7.2%. On the other, household without married couple is 14%.

Number of	Number			Number of	Percentage
couples'	of	Household consist of the below couples	Code2	households	(%)
generation	couples			nousenoius	(70)
		All households		9,189	100.0
		All households with marriage couples and without other		7,445	81.0
		relation			
1 generation		Total		6,774	73.7
	1	Total		6,769	73.7
		♦ Subtotal of a head's couple	10	6,350	69.1
		♦ Nuclear family (Sum of 11 and 12)		5,555	60.5
		Head's couple only	11	570	6.2
		Head's couple and their children	12	4,985	54.2
		Head's couple, their children and their grandchildren	13	177	1.9
		Head's couple and their parents	14	26	0.3
		Head's couple, their parents and their children	15	488	5.3
		Head's couple, their parents, their children and their	16	15	0.2
		grandchildren			

Table Type of household by number of married couples' generation and number of married couples

	head's couple, their grandparents and their parents	17	0	-
	head's couple, their grandparents, their parents and	18	1	0.0
	their children			
	Head's couple and their grandchildren	19	57	0.6
	Head's couple, their grandparents and their children	20	24	0.3
	Head's couple, their parents and their grandchildren	21	3	0.0
	Other	29	4	0.0
	♦ Subtotal of no head's couple	30	419	4.6
	No head's couple but children's couple	31	400	4.4
	No head's couple but parents' couple	32	19	0.2
	No head's couple but siblings' couple	33	0	-
	No head's couple but grandchildren's gouple	34	0	-
2 and	Total	40	5	0.1
more				
	No head's couple but children's couples	41	5	0.1
	Head's couple and siblings' couples	42	0	-
	No head's couple but siblings' couple	43	0	-
	Total	50	664	7.2
	Head's couple and children's couple	51	598	6.5
	Head's couple and parents' couple	52	64	0.7
	Head's couple and grandchildren's couple	53	0	-
	Head's couple and grandparents' couple	54	2	0.0
	Head's couple, parents' couple and children's couple	60	7	0.1
	No married couple without other relation	70	1,262	13.7
	Households with other relations	80	482	5.2
	2 and more	head's couple, their grandparents and their parents head's couple, their grandparents, their parents and their children Head's couple and their grandchildren Head's couple, their grandparents and their children Head's couple, their grandparents and their children Head's couple, their parents and their grandchildren Other Other No head's couple but children's couple No head's couple but parents' couple No head's couple but grandchildren's gouple No head's couple but grandchildren's gouple 2 and Total Mead's couple but children's couples Head's couple but siblings' couples No head's couple but siblings' couples No head's couple but siblings' couples Head's couple and siblings' couples Head's couple and children's couple Total Head's couple and parents' couple Head's couple and parents' couple Head's couple and grandparents' couple	head's couple, their grandparents and their parents 17 head's couple, their grandparents, their parents and their children 18 Head's couple and their grandchildren 19 Head's couple, their grandparents and their children 20 Head's couple, their grandparents and their children 21 Other 29 ◇ Subtotal of no head's couple 30 No head's couple but children's couple 31 No head's couple but parents' couple 32 No head's couple but grandchildren's gouple 34 2 and Total 40 more No head's couple but children's couples 41 Head's couple but siblings' couples 42 No head's couple but children's couples 41 Head's couple but siblings' couples 42 No head's couple but siblings' couples 42 No head's couple and siblings' couple 50 Head's couple and parents' couple 51 Head's couple and parents' couple 52 Head's couple and grandparents' couple 53 Head's couple and grandparents' couple 54 Head's couple and grandparents' couple 54 <tr< td=""><td>head's couple, their grandparents and their parents170head's couple, their grandparents, their parents and their children181Head's couple and their grandparents and their children1957Head's couple, their grandparents and their children2024Head's couple, their parents and their grandchildren213Other294\diamond Subtotal of no head's couple30419No head's couple but children's couple31400No head's couple but parents' couple3219No head's couple but grandchildren's gouple3402 and moreTotal4055Mo head's couple but children's couples415Head's couple but children's couples415Head's couple but children's couples420No head's couple but children's couples420No head's couple but children's couples4302 and MoreTotal50664Head's couple and siblings' couple51598Head's couple and parents' couple5264Head's couple and grandchildren's couple530Head's couple and grandchildren's couple542Head's couple and grandparents' couple542Head's couple and grandparents' couple607No married couple without other relation701,262Households with other relations80482</td></tr<>	head's couple, their grandparents and their parents170head's couple, their grandparents, their parents and their children181Head's couple and their grandparents and their children1957Head's couple, their grandparents and their children2024Head's couple, their parents and their grandchildren213Other294 \diamond Subtotal of no head's couple30419No head's couple but children's couple31400No head's couple but parents' couple3219No head's couple but grandchildren's gouple3402 and moreTotal4055Mo head's couple but children's couples415Head's couple but children's couples415Head's couple but children's couples420No head's couple but children's couples420No head's couple but children's couples4302 and MoreTotal50664Head's couple and siblings' couple51598Head's couple and parents' couple5264Head's couple and grandchildren's couple530Head's couple and grandchildren's couple542Head's couple and grandparents' couple542Head's couple and grandparents' couple607No married couple without other relation701,262Households with other relations80482

- > table(ft2\$p7>0)
- FALSE TRUE
- 8707 482
- > table(ft2\$p7==0&ft2\$ct==0)
- FALSE TRUE
- 7927 1262
- > d<-subset(ft2,p7==0&ct>0)
- > dim(d)
- [1] 7445 18

> d["code2"]<-ifelse(d\$c1==0&d\$c2>0&d\$c4>0&d\$c5==0, 60, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==0&d\$c3>0&d\$c4>0&d\$c5==0,51,d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2>0&d\$c3>0&d\$c4==0&d\$c5==0, 52, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==0&d\$c3>0&d\$c4==0&d\$c5>0,53,d\$code2) > d["code2"]<-ifelse(d\$c1>0&d\$c2==0&d\$c3>0&d\$c4==0&d\$c5==0, 54, d\$code2) > d["code2"] <-ifelse(d\$ct==1, 1, d\$code2)> d["code2"]<-ifelse(d\$ct==1&d\$c3==1, 29, d\$code2)</pre> > d["code2"]<-ifelse(d\$ct==1&d\$c3==0, 30, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==0&d\$c3==0&d\$c4==1&d\$c5==0&d\$c6==0, 31, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==1&d\$c3==0&d\$c4==0&d\$c5==0&d\$c6==0, 32, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==0&d\$c3==0&d\$c4==0&d\$c5==0&d\$c6==1, 33, d\$code2) > d["code2"]<-ifelse(d\$c1==0&d\$c2==0&d\$c3==0&d\$c4==0&d\$c5==1&d\$c6==0, 34, d\$code2) > d["code2"]<-ifelse(d\$ct>=2&d\$c1==0&d\$c2==0&d\$c3==0&d\$c4>=2&d\$c5==0&d\$c6==0, 41, d\$code2) > d["code2"]<-ifelse(d\$ct>=2&d\$c1==0&d\$c2==0&d\$c3==1&d\$c4>=2&d\$c5==0&d\$c6>=1, 42, d\$code2) > d["code2"]<-ifelse(d\$ct>=2&d\$c1==0&d\$c2==0&d\$c3==0&d\$c4>=2&d\$c5==0&d\$c6>=2, 43, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$hhsz==2, 11, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2==0&d\$p4>0&d\$p5==0&d\$p6==0, 12, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2==0&d\$p4>0&d\$p5>0&d\$p6==0, 13, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2>0&d\$p4==0&d\$p5==0&d\$p6==0, 14, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2>0&d\$p4>0&d\$p5==0&d\$p6==0, 15, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2>0&d\$p4>0&d\$p5>0&d\$p6==0, 16, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1>0&d\$p2>0&d\$p4==0&d\$p5==0&d\$p6==0, 17, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1>0&d\$p2>0&d\$p4>0&d\$p5==0&d\$p6==0, 18, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2==0&d\$p4==0&d\$p5>0&d\$p6==0, 19, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1>0&d\$p2==0&d\$p4>0&d\$p5==0&d\$p6==0, 20, d\$code2) > d["code2"]<-ifelse(d\$ct==1&d\$c3==1&d\$p1==0&d\$p2>0&d\$p5>0&d\$p6==0, 21, d\$code2) > table(d\$code2) 11 12 13 14 15 16 18 19 20 21 29 31 32 41 51 570 4985 177 26 488 15 24 3 400 19 5 598 1 57 4 52 54 60 2 7 64 > round(table(d\$code2)/nrow(ft2)*100, 1) 14 15 16 18 19 20 21 29 31 32 41 51 11 12 13 6.2 54.2 1.9 0.3 5.3 0.2 0.0 0.6 0.3 0.0 0.0 4.4 0.2 0.1 6.5 52 54 60

> d["code2"]<-0

 $0.\ 7 \quad 0.\ 0 \quad 0.\ 1$

Added code2 to ft2

> ft2<-merge(ft2,d[c("ID", "code2")],by="ID",all=T)

> ft2["code2"]<-ifelse(ft2\$p7>0,80,ft2\$code2)

> ft2["code2"]<-ifelse(ft2\$p7==0&ft2\$ct==0,70,ft2\$code2)

> table(ft2\$code2, useNA="ifany")

11 12 13 14 15 16 18 19 20 21 29 31 32 41 51 570 4985 177 26 488 15 1 57 24 3 4 400 19 5 598 52 54 60 70 80 64 2 7 1262 482 > ft2.old3<-ft2

Family type by region

Comparing family type between Red River Delta and Mekong River Delta, the share of nuclear family is higher in Red River Delta. Consequently, the average household size in Red River Delta is smaller than Mekong River Delta.

 \checkmark Composition of type of households by number of household members' generation and by region

> family<-cbind(ft2,ttchung[,c("tinh","ttnt","dantoc","tsnguoi","thunhap","chids","wt")])

> t<-tapply(family\$wt, list(substr(family\$tinh, 1, 1), family\$code), sum)

> t[is.na(t)]<-0

> region.name<-c("Red River Delta", "North East", "North West", "North Central",

+ "South Central Coast", "Central Highlands", "South East", "Mekong River Delta")

> rownames(t) <-region.name

> round (prop. table(t, 1)*100, 1)

	11	12	21	22	31	32	33	35	41	42	43	44	50	60
Red River Delta	4. 6	10. 3	62. 0	0.7	10. 6	1.3	6. 9	0. 0	0.5	0.1	0.0	0.1	0.0	2. 9
North East	1.9	5.1	64. 8	0.9	10. 0	0.6	9.3	0. 0	0.5	0.0	0.1	0.6	0.1	6. 1
North West	1.6	3.6	66. 6	0.6	10. 8	0.4	5.5	0. 0	0.5	0.0	0.0	0.0	0.0	10. 4
North Central	4. 1	7.3	65.9	0.5	10. 4	0.4	7.6	0. 1	0. 7	0.0	0.0	0.4	0.0	2. 6
South Central Coast	3. 5	6.4	65. 2	0.6	10. 5	0. 7	8. 4	0. 1	0.4	0.0	0.0	0.3	0.0	3.9
Central Highlands	1.9	4. 1	72. 0	0. 2	7.8	0. 7	5.9	0. 0	0.4	0.0	0.0	1.0	0.0	5.8
South East	5.5	3.5	61.8	0.7	13.6	1.0	5.0	0.0	0.6	0.0	0.0	0.3	0.0	8.0
Mekong River Delta	4. 0	5.2	59.7	1.3	16.5	1.8	4.4	0. 0	0.3	0. 1	0.0	0.1	0.0	6.4

✓ Composition of type of households by number of married couples' generation and number of married couples, and by region

> t<-tapply(family\$wt, list(substr(family\$tinh, 1, 1), family\$code2), sum)

> t[is.na(t)]<-0

> round (prop. table(t, 1)*100, 1)

 11
 12
 13
 14
 15
 16
 18
 19
 20
 21
 29
 31
 32
 41
 51
 52
 54
 60
 70
 80

 1
 10.3
 52.9
 1.8
 0.4
 5.6
 0.2
 0
 0.9
 0.1
 0.1
 0.0
 4.1
 0.0
 0.0
 5.4
 0.7
 0.0
 0.1
 14.6
 2.9

 2
 5.1
 55.5
 1.5
 0.7
 7.5
 0.1
 0
 0.5
 0.4
 0.0
 0.1
 3.3
 0.1
 0.0
 7.7
 1.1
 0.0
 0.1
 10.2
 6.1

 3
 3.6
 57.7
 1.8
 0.6
 3.9
 0.0
 0
 0.4
 0.0
 0.1
 1.8
 0.0
 0.0
 10.4
 1.0
 0.0
 0.2
 8.2
 10.4

 4
 7.3
 56.7
 2.2
 0.3
 6.7
 0.3
 0
 0.2
 0.4
 0.0
 0.1
 3.7
 0.2
 0.1
 5.1
 0.4
 0.0
 0.0
 13.3
 3.9

8	5.2 50.6 2.3 0.2 3.5 0.0	0 0.9 0.0 0.1 0.0 6.3 0.4 0.1	8.70.70.10.114.5	6.4
7	3.5 50.2 2.3 0.1 3.7 0.0	0 0.4 0.3 0.0 0.0 6.6 0.2 0.1	5.60.40.00.018.5	8.0
6	4. 1 62. 8 1. 9 0. 2 5. 1 0. 2	0 0.5 0.7 0.0 0.1 2.1 0.0 0.0	4. 2 0. 6 0. 2 0. 3 11. 2	5.8

\checkmark Average household size by region

> hhsz<-tapply(family\$tsnguoi*family\$wt,substr(family\$tinh,1,1),sum)/

+ tapply(family\$wt,substr(family\$tinh,1,1),sum)

> names(hhsz) <-region.name

> round (hhsz, 2)

Red River Delta	North East	North West	North Central
3. 83	4. 27	4.81	4. 27
South Central Coast	Central Highlands	South East	Mekong River Delta
4. 27	4. 88	4. 31	4. 19

6.8 Education

- Compared school attendance rate (SRA) between male and female.
- Data frames and variables used;
 - \blacktriangleright mucla 39,071 records
 - ♦ MIAC2 sex
 - ♦ M1AC5 age
 - ♦ wt
 - ♦ PID
 - ➤ muc2a 39,071 records
 - ♦ M2AC5 Currently attending school?
 - 1 Yes
 - 2 Vacation
 - 3 No
 - ♦ wt
 - ♦ PID

```
> d<-merge(muc1a[, c("PID", "m1ac2", "m1ac5")], muc2a[, c("PID", "m2ac5", "wt")])
> d<-subset(d,m1ac5>=5&m1ac5<=24)
> dim(d)
[1] 15835
              5
> m<-matrix (as. vector (by (d$wt, list (d$m1ac5, d$m2ac5, d$m1ac2), sum)), nrow=20, ncol=6)
> rownames (m) < -5:24
> colnames (m) <-outer (c("Yes", "Vacation", "No"), c("M", "F"), paste, sep="-")
> round (m)
    Yes-M Vacation-M
                       No-M Yes-F Vacation-F
                                                  No-F
5
  256176
              126693 196230 227789
                                         98104 183635
  280540
6
              206955 102211 298711
                                        172225 60424
  306948
              279764 17811 307781
7
                                        180568
                                                  8203
8
  300101
              283586 14080 306068
                                        262342 21313
9
  357408
              332475 18410 358804
                                        304737 17060
10 452439
              394041 20013 400681
                                        365906
                                                  9140
11 382831
              354612 37193 475993
                                        386252 28489
12 420381
              367452 40795 483817
                                        406042 52599
13 446986
              460903 94640 498525
                                        462536 86768
14 457009
              387374 159974 438114
                                        418141 141029
15 455399
              413641 258209 365051
                                        384540 202645
16 445887
              335297 376000 412619
                                        345752 329362
17 335312
              269601 483310 348163
                                        272138 360918
18 271186
              155011 582531 207864
                                        139435 492442
19 270177
               46581 645679 277946
                                         40373 504163
20 226579
               39338 576048 215340
                                         33407 535029
```

21 185678 22555 641390 189922 22000 575593 22 154610 11020 638537 132590 9215 572484 23 142607 17782 617486 55919 4953 679878 24 55446 1604 647971 48171 10649 603295 # School attendance rate by sex and single year age > df<-as. data. frame(m) > df["male.SAR"]<-(df[,1]+df[,2])/rowSums(df[,1:3])*100 > df["female.SAR"]<-(df[,4]+df[,5])/rowSums(df[,4:6])*100 > (sar<-round(df[, 7:8], 1)) male.SAR female.SAR 5 66.1 64.0 82.7 6 88.6 7 97.1 98.3 8 97.6 96.4 9 97.4 97.5 10 97.7 98.8 11 95.2 96.8 12 95.1 94.4 13 90.6 91.7 14 84.1 85.9 15 77.1 78.7 16 67.5 69.7 17 55.6 63.2 42.3 18 41.4 19 32.9 38.7 20 31.6 31.7 21 24.5 26.9 22 20.6 19.9 23 8.2 20.6

> write.table(sar,file="clipboard",sep="\t",row.names=F)

8.9

✓ Summary:

8.1

24

Female's school attendance rate (SAR) is slightly higher than male's SAR.



7. Agricultural land

7.1 History

- In Vietnam, land belongs to the entire people, which was approved by the National Assembly in 1980. The State has the powers and responsibilities for uniform administration of land as representative.
- In 1986, Doi Moi (Renovation) policy was launched.
- 1993 Land Law allowed ownership of a right to use land. This right is called the Land Use Right (LUR).
- In 1993, agricultural land was allocated free of charge for farmers for 20 years.
- 2003 Amended Land Law introduced market mechanism for LUR. Market price was applied for the price of LUR.

7.2 Rice cultivated area by region

Comparison of the rice cultivated area by region shows that mean rice cultivated area per sample household conditional on non-zero rice production is smaller in Red River Delta as compared to Mekong River Delta. The coefficient of variation is also smaller in Red River Delta.

It is said that in North Vietnam agricultural land was equally distributed among farmers after the World War II, which resulted in the small mean area and the coefficient of variation. On the other, in South Vietnam landlord system had continued until the unity in 1976.

Table	Share of sample	households	cultivated	rice,	mean	and	coefficient	of	variation	of	rice
cultivat	ted area per samp	le household	(VHLSS 2	2006)							

Region	Share of households	Mean rice cultivated area	Coefficient of	
	cultivated rice (%)	for the past 12 months (m ²)	variation	
Red River Delta	67.6	4,019	0.63	
Mekong River Delta	37.2	23,141	1.12	

Note: For example, in a period of one year, on a land lot of 360 m^2 people grow two rice crops, total rice cultivated area is twice the arable area, that is 720 m^2 .

Used data frame MUC4B11

> d<-subset(muc4b11, m4b11ma>=5)

> dim(d)

[1] 6477 22

Aggregated the rice cultivated area for the past 12 months by household

> t<-tapply(d\$m4b11c3, d\$ID, sum)

Created data frame land consist of variables of ID, area, wt for all households.

> land<-data.frame(ID=names(t), area=t, row.names=NULL)

> land<-merge(land, ttchung[c("ID", "wt")], all=T)</pre>

> land[is.na(land)]<-0

Mean rice cultivated area by region

> tapply(land\$area, substr(land\$ID, 1, 1), function(x) mean(x))

1 2 3 4 5 6 7 8

2718.764 2871.756 5305.886 3175.130 2344.995 2342.837 1607.811 8620.547 # Coefficient of variation

> tapply(land\$area, substr(land\$ID, 1, 1), function(x) sqrt(var(x))/mean(x))

 1
 2
 3
 4
 5
 6
 7
 8

 1.029775
 1.047827
 1.154190
 1.316736
 1.316790
 2.056574
 3.938203
 2.248628

Generated data frame land2, a subset of land, where household cultivated rice.

> land2<-subset(land, area>0)

> dim(land2)

[1] 4824 3

Share of sample households cultivated rice

> tapply(land2\$wt, substr(land2\$ID, 1, 1), length)/tapply(land\$wt, substr(land\$ID, 1, 1), length) 1 2 3 4 5 6 7 8 0. 6764403 0. 7031131 0. 7715618 0. 6607495 0. 5516432 0. 4106529 0. 1506734 0. 3725174 # Mean rice cultivated area per sample household > tapply(land2\$area, substr(land2\$ID, 1, 1), function(x) mean(x)) 1 2 3 4 5 6 7 8 4019. 223 4084. 343 6876. 813 4805. 346 4250. 928 5705. 151 10670. 838 23141. 324 # Coefficient of variation

> tapply(land2\$area, substr(land2\$ID, 1, 1), function(x) sqrt(var(x))/mean(x)) 1 2 3 4 5 6 7

0. 6274495 0. 6892151 0. 8941031 0. 8980131 0. 7128223 1. 0720696 1. 2222574 1. 1211101

8