Chapter 3

Sampling and the Analysis of Data

3.1 Research Design

Stratified sampling method is used to examine the income, capabilities and social capital of households of Chinese in Chiang Mai. The households sampling are divided into four groups: 1. Random sample Thai people living in Chiang Mai 2-19 years; 2. Random sample Thai people living in Chiang Mai more than 20 years; 3. Random sample Chinese living in Chiang Mai 2-19 years; 4. Random sample of Chinese living in Chiang Mai here more than 20 years. In order to research the comprehensive effects of information of sample, the number of each group is 50 households respectively. But if most Chinese people living in Chiang Mai fled China for political reasons during the period 1947-1950, i.e. they are all long-term migrants, the whole sampling frame and sample size will go down.

The samples are all from other places and came to Chiang Mai with the family. The native Thai sample came from other provinces of Thailand like Bangkok and Chiang Rai. And the source of Chinese sample is migrants who were from Yunnan, Guangxi, Fujian, Shanghai and so on.

For knowing the overall comparative income, capabilities and social capital of Chinese and Thai households in Chiang Mai, both subjective and objective measurements of well being are used to investigate basic family information,

physical wellbeing, emotion wellbeing, capabilities and social capital.

The location of research is Chiang Mai, which is the largest city in north of Thailand. It's near to China and absorbs lots of Chinese migrate with the whole family. Based on the key finding of the survey, theory thesis is conducted.

The survey is used for data collection in the months of January-February 2011, with the key information got mainly from the head of households. The secondary data from government is also selected.

The questionnaire is designed to reflect both exterior and interior characteristic. The exterior questions (like the size of dwelling) are asked and checked by interviewer. And most sensitive questions (like income and inner world) are set to choose the range or level, so that people can reasonably be expected to ask. In this research, the data is mainly measured by number (such as wealth) and seven-point Likert scale (such as capabilities, social capital, and happiness). The questions are divided into the following parts:

Basic household information – This part is mainly examined by the size of households, religion, length of stay in Chiang Mai, reason for coming to Chiang Mai of households. And the gender, age, educational attainment, marital status of the member of household are also considered.

Health status – Health status is the body diathesis of households and the situation of sickness, disease and medical consultation.

Dwelling – Dwelling is analyzed by basic life facilities, such as gas,

television, and internet and so on. The size and type of dwelling are also considered in this part.

Income and Expenditure – This section is mainly emphasized on occupation, the range of income, expenditure per month and the purpose of expenditure. The income is in the form of salary and cash income received from other sources (like sale of valuables, rent, alimony or charity).

Benefits – Pension, disability, child benefit, and unemployment benefit are involved in this part.

Savings, Assets and Loans – These indicators reflect the wealth of households.

Inequality – This part test the thinking of households of social inequality. The households could compare themselves with other Chinese households in Chiang Mai, native Thai households in Chiang Mai and Chinese households in China. The responders choose the main elements of inequality and the way of decreasing the gap between rich and poor people.

Emotional wellbeing – This indicator is conducted as the subjective measurement by asking the questions of how satisfied with their lives, work and leisure life; the happiness of their mind, body, heart, soul and with others. Capabilities and Sufficiency are also concluded as subjective measurement by choosing the level of several indicators, such as rights, freedoms, safety, and the achievement of willingness.

Anther important indicator is social capital. It's a test of households' network of family, friends, colleagues in Chiang Mai and the connection with hometown. The research analyzes the bonding capital, bridging capital, social capital per capita and total household social capital. The bonding capital (horizontal capital) measures the network of members relates to each other on an equal basis, and the bridging capital (vertical capital) tests the network of hierarchical relationships and unequal power distribution among members. The social capital per capita, as a noble idea of this research, is the average social capital for each person in every household. And the total social capital is defined as the social capital per household.

3.2 Conceptual Framework of the Study

In order to structure the components of this thesis, the framework of the research was shown like following. The framework indicates the main research factors and the relationship among them.

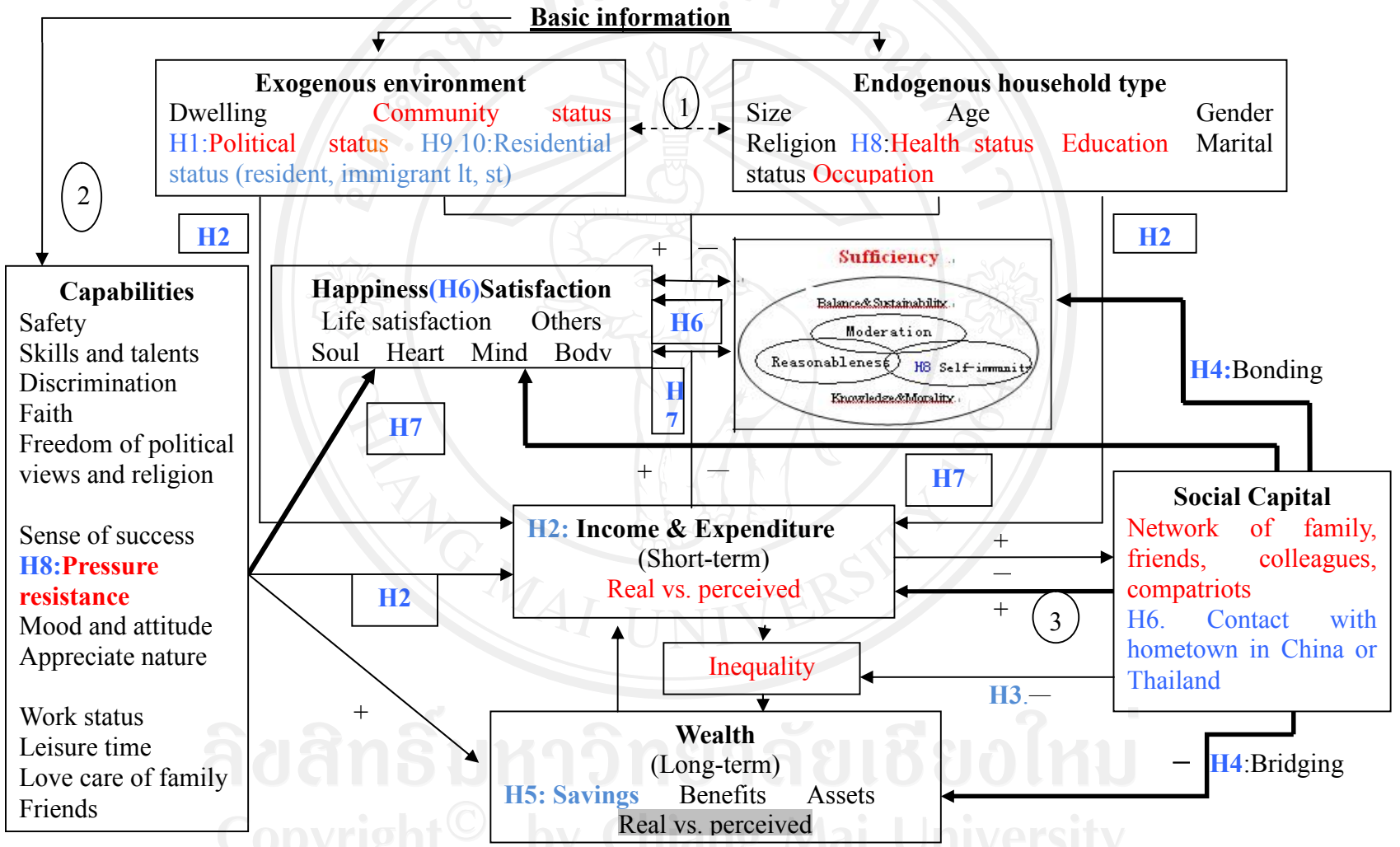

Figure 3.1: The framework of the research

In *horizontal* (in the Chinese or Thai households group) area, the research starts from the basic information of Chinese and Thai households in Chiang Mai. The information will be got from survey by asking the endogenous (such as household size, age, and gender) and exogenous (such as dwelling, political and residential status) questions. There maybe have some weak relationship between these two elements (arrow 1).

The first main factor is income. It's a relative short-term indicator of households' wealth. The Gini-coefficient and Theil index will be used for revealing the inequality among Chinese and Thai people in Chiang Mai. Besides, the influencing factors of income –basic information, capabilities and social capabilities (strong and positive) - are showed by arrow H2 and arrow 3. This research will investigate the difference between real income (wealth) and perceived income (wealth). From the question of feelings of equality, the perceived equality will be get.

The second main factor in this research is capabilities. It involves several elements like freedom, mood and social status. The effect of basic information (arrow 2) will be researched in this study as well.

The last main factor is social capital of households. It mainly analyzes the network of family, friends, colleagues and contact with hometown. From hypothesis 3, the social capital may have negative effect on inequality (arrow H3) and households' wealth (arrow H4), however, is affected by income in positive or

negative way. The hypothesis 4 explores the strong bonding effect of social capital in Chinese households and the strong bridging effect of social capital in Thai households (arrow H4).

The satisfaction and happiness are also measured in this study. From hypothesis 6, the satisfaction has relationship with sufficiency which is analyzed by a system. Arrow H7 shows that the happiness is affected by capabilities and social capital more than by income. And the basic information may also have positive and negative relationship with happiness and sufficiency.

In *vertical* (between Chinese and Thai households groups) area, the following indicators may quite different between Chinese and Thai households from the hypothesis: political status, education, pressure resistance, inequality, life satisfaction and social capital. The research will compare Chinese households with native Thai households to check whether the difference of these indicators is significant or not.

3.3 Methods of Data Analysis

After selecting and checking the data, we can analysis them. On one hand, the *horizontal* data (in the Chinese or Thai group) is focused on the analysis of the factor that impact on income, capability, social capital or happiness. The relationship among main indicators is also analyzed. On the other hand, the *vertical* data (between Chinese and Thai group) explains the difference in indicators between Chinese and Thai group.

Lorenz curves and Gini-coefficient are used to explain the income distribution of each group. Other analytical tools like t-test and regression are also processed by SPSS 17.0.

3.3.1 Descriptive Statistics

Descriptive statistics are involved in a table processed by SPSS 17.0. We can get the basic information of every variable. Descriptive statistics include the number of observed values, minimum and maximum values, means and standard deviation. Standard deviation divided by mean gets the coefficient of variation.

The number of observed values, minimum and maximum values and means can tell the primary and general information of variables. At the same time, they can help us to check the data of survey. The coefficient of variation gives a way to judging weather the variable is valuable to analyze or not. If the coefficient of variation of a variable is bigger than or equal to 200%, it's a useful factor for this research. If this number of a variable is less than 30%, it may be not a valuable factor.

3.3.2 Analysis of Income

The analysis of income used in this research is Gini-coefficient and Theil index. The Lorenz Curve will also show the income distribution.

3.3.2.1 Lorenz Curve

The Lorenz Curve is a way of illustrating the income distribution. It shows the actual and perfectly equal distribution of income. The horizontal axis shows the

cumulative share of population from lower income. And the vertical axis shows the cumulative share of income.

The perfect equality line is a 45 degree line starting from original point. It stands for x% of the population having x% of the income. The more curved the actual income line the more unequal the distribution of income. The graphical representation of Lorenz curve is liking following:

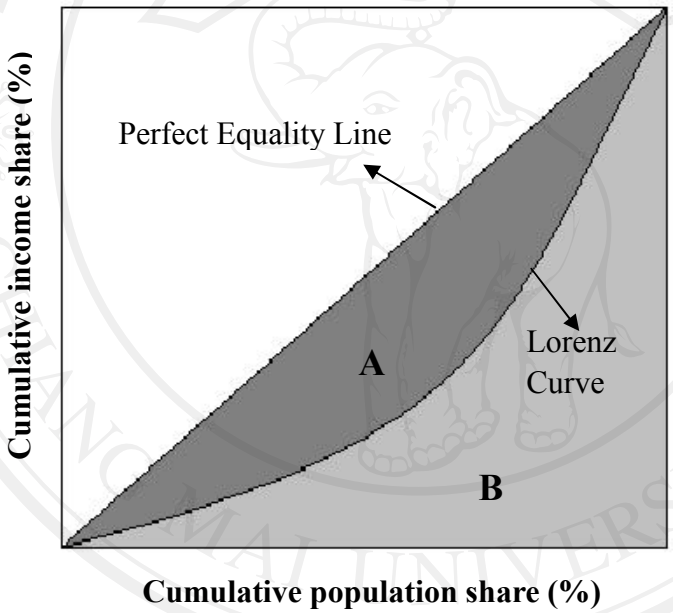

Figure 3.2: Lorenz Curve and Gini Coefficient

3.3.2.2 Gini-coefficient

The Gini-coefficient is defined as a ratio of the areas on the Lorenz curve diagram. Denoting the Gini-coefficient by G, then

$$G = \frac{A}{A+B} \qquad (A+B = 0.5)$$

From figure 3.2, A is the area between the Lorenz curve and the perfect equality distribution, and B is the area under the Lorenz curve. When there is total equality the Lorenz curve coincides with the 45°line, area A then disappears and G = 0. With inequality, area disappears and G = 1. Thus, G is between 0 and 1 and the smaller of G (area A is smaller) the more equal of distribution.

The main advantage of Gini-coefficient is that it is a measure of inequality by means of a ratio analysis, rather than a variable unrepresentative of most of the population. And it can be used to compare income distributions across different population sectors as well as countries. We can use Gini-coefficient to compare the inequality between urban and rural area in a country.

The Gini-coefficient and Lorenz curve have some advantages[1]:

1. The Lorenz curve may understate the actual amount of inequality if richer households are able to use income more efficiently than lower income households. From another point of view, measured inequality may be the result of more or less efficient use of household incomes.
2. Economies with similar incomes and Gini-coefficients can still have very different income distributions. This is because the Lorenz curves can have different shapes and yet still yield the same Gini-coefficient. As an extreme example, an economy where half the households have no income and the other half share income equally has a Gini-coefficient of 1/2; but

[1] From Dr. Schoch, D.'s lecture- Welfare Economics.

an economy with complete income equality, except for one wealthy household that has half the total income, also has a Gini-coefficient of 1/2. In practice, such distributions do not exist, and therefore, the impact of different but realistic curves is less obvious.

3. The Gini-coefficient of different sets of people cannot be averaged or otherwise aggregated to obtain the Gini-coefficient of all the people in the sets: if a Gini-coefficient were to be calculated for each person it would always be zero. We can use decomposable inequality measures such as Theil index to solve this problem.

The main calculation of Gini-coefficient and the map of Lorenz curves are processed by Excel in this research.

3.3.2.3 Theil Index

Theil index of this research defined by

$$T_1 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\bar{x}} \bullet \ln \frac{x_i}{\bar{x}}\right)$$

Where, N= the number of individual or household, x_i= the income of ith individual or household, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the average income. The first term inside the sum is the individual's share of aggregate income, and the second term is the person's income relative to mean. If everyone is equal, the index is 0. If one person has all the income, then the index is lnN.

To consider the weighted average of inequality within and among subgroups, we suppose that the population is divided into m subgroups,

$$T_2 = \sum_{k=1}^{m} s_k T_k + \sum_{k=1}^{m} s_k \ln \frac{\overline{x}_k}{\overline{x}}$$

Where s_k=the income share of group k, T_k=the Theil index for subgroup k, $\overline{x}_k$ =the average income in group k. The first part of equation is a weighted average of inequality within subgroups, and the second part is the inequality among those subgroups.

3.3.3 Analysis of Other Indicators

The other indicators such as capability, social capital and happiness are analyzed mainly by SPSS 17.0.

3.3.3.1 Frequency

The analysis of frequency includes the frequency, percent, valid percent and cumulative percent. The frequency is a measure of the number of occurrences of a repeating event per unit of time. In this research, it's used for measure if major of Chinese migrant came to Chiang Mai between 1947 to1950 (for about 61 or 64 years).

3.3.3.2 Crosstabs

The crosstabs is the process of creating a contingency table from the multivariate frequency distribution of statistical variables. It is a table that we can analyze the components of variables by different group. In this study, it tells the

compared percentage of different immigration motivation by two ethnic groups. Besides this, crosstabs show the bar chart of motivation of immigration by two groups.

3.3.3.3 Correlation

Correlation is one of most commonly used statistics in research. A correlation is a number that describes the degree of relationship (positive or negative) between two variables. It can help us to find out multicollinearity and highly associated variables. Two variables can not be independent variables used in one regression equation to explain a dependent variable if they have high correlation. The level of significance (2-tailed) of correlation divides into less than 0.1, 0.05 and 0.01. It is common practice to mark these three levels with *, **, and *** respectively.

In this research, correlation is used to find out the associated factors of a main variable. If the significance of variable A and B is less than 0.1 or Pearson Correlation is more than or equal to 0.5, the correlation is significant, and we can say B associate A. If the correlation of A and C is not significant, C is not associated A. Then B rather than C can be an independent variable in the regression equation of dependent variable A. But A and B can not be independent variables used in another regression equation to explain dependent variable, let's say, X.

3.3.3.4 T-statistics

T-test in this research mainly used to discuss whether the difference between two variables or two ethnic groups is significant or not. The T-test statistic must be

greater than or equal to 2.58 for the 99% level of confidence interval, 1.96 for the 95% level, or 1.67 for the 90% level. In the result of SPSS, T-test is included in the independent samples test table or the one-sample statistics table. If the significance (2-tailed) is less than 0.01(***), 0.05 (**) and 0.1 (*) respectively, the difference in the tested variable between two ethnic groups is significant.

3.3.3.5 F Statistics

F statistics is used to indicate the degree of explanation of a model. The model is acceptable when the F statistic is higher than 4 and significance is less than 0.1.

3.3.3.6 One-way ANOVA

We use one-way ANOVA to avoid multiple T-tests and decrease the opportunity to get the wrong results. In this research, the analysis of ANOVA aims to get the relationship between two variables which include several indicators. The confidence interval is 95% with SPSS 17.0. So if the p value in multiple comparisons table is less than 0.05, the variables are different.

3.3.3.7 Regression

From the results of regression, we can know the model for a hypothesized variable. It is the relationship between dependent variable and independent variables. There are one dependent variable and one or more independent variables in the regression equation. The beta coefficients are the degree of impact of independent variables on dependent variable. If the beta coefficient is positive, the independent

variables are directly related to the dependent variable and if it is negative, the independent variables are inversely related to the dependent variable. The constant indicates the starting point of regression line on the vertical axis. If constant is equal to 0, the regression line starts from the original point.

3.4 Variables

The variables are got from the survey and mainly analyzed by SPSS 17.0. There are 200 households (100 Thai households and 100 Chinese households) with 329 variables for each household.

The measurement of variables is separated into three groups: cardinal, ordinal and nominal. The cardinal variables are interval numbers. For example, the time in Chiang Mai is a cardinal variable. Some of the cardinal variables could be set the values. For instance, the gender of household head is male if the value is 0 and female if the value is 1. Ordinal variables are variables whose order is significant. In this research, the ordinal variables are all about frequency and level. The scale of the ordinal variables is from 1 to 7 (or 6, 12). The variable is nominally measurable if it's not a numerical value, such as the birthplace.

The variables in this research could be separated in several parts:

Table 3.1: The Definition of Related Variables in This Research

Category	Variable Name	Description	Measure
Income	a. Income per capita	Average income for each person in every household.	Ordinal
	b. Total Income	Total income of each household.	

Table 3.1. (Continued)

Category	Variable Name	Description	Measure
Capability	a. Bodily Health	The health limitation of interviewee.	Ordinal
	b. Bodily Integrity	The security of interviewee.	
	c. Senses Imagination & Thought	Usage and freedom of imagination and thought in work and life.	
	d. Emotions	Feeling of friendships, support, love, anger and strain.	
	e. Practical Reason	The capacity of judging and controlling life.	
	f. Affiliation	Being able to living with others and the importance for other people; Respect to other objects and self-respect.	
	g. Other Species	Concern for plants, animals and the world of nature.	
	h. Play	Leisure activities.	
	i. Control over One's Environment	Respect to and importance of other people in work.	
Social Capital	a. Bonding Capital	Network of members on an equal basis.	Ordinal
	b. Bridging Capital	Network of members that at different levels or have unequal power.	

Table 3.1. (Continued)

Category	Variable Name	Description	Measure
	c. Total Social Capital	Total social capital of each household.	
	d. Social Capital per capita	Average social capital for each person in every household.	
	e. Number of Participated Organizations	The number and name of participated organizations.	Cardinal
Satisfaction	a. Life Satisfaction	Satisfaction of economic, living and working conditions, interpersonal relations, community environment, children's education and leisure life.	Ordinal
Happiness	a. Happiness of Others	The feeling of others' activities; Love to share with others.	Ordinal
	b. Happiness of Soul	Peace in soul; Accept others' different beliefs.	
	c. Happiness of Heart	Life attitude; The feeling of shame, guilt and blame.	
	d. Happiness of Mind	Freedom and peace in mind; Control mind in life.	
	e. Happiness of Body	Attitude in face of difficulty.	
	f. Happiness of Volunteering	Feeling of non-profit activities.	

Table 3.1. (Continued)

Category	Variable Name	Description	Measure
Education	a. Education of household head	Separated into illiterate, literate but without formal schooling, primary, secondary, university and professional degree.	Ordinal
	b. Highest Education		
	c. Average Education		
Wealth	a. Saving, assets and benefits	The total of savings, assets and benefits of each household per month.	Cardinal
Demand of Life Overseas	a. Perceived Income	Minimum income for living well.	Cardinal
Socio-demography	a. Gender of Household Head	Separated into male (value =0) and female (value =1). Expressed as female household head.	Cardinal
	b. Number of Family Members	Analyze the number of total family members, males and females in family.	
	c. Age of Family Members	Analyze the average age of family members and the age of household head.	
	d. Marital Status of Household Head and Interviewees	Separated into married, co-habiting, never married, divorced, separated and widowed.	Ordinal

Table 3.1. (Continued)

Category	Variable Name	Description	Measure
Other Variables	a. Ethnic Group	Separated into Thai households (value =0) and Chinese households (value =1). Expressed as Chinese households.	Cardinal
	b. Time in Chiang Mai	Length of stay in Chiang Mai.	
	c. Motivation of immigration	The reason of migration to Chiang Mai, for occupation, business or political reason.	Ordinal
	d. Sufficiency	Knowledge and usage of Sufficiency Economy Philosophy.	
	e. Inequality	Feeling of inequality with other Thai/Chinese people in hometown/Chiang Mai.	

3.5 Questionnaire

The questionnaire of this study is designed to inquire about the income, capability and social capital of Thai and Chinese households in Chiang Mai. The interviewee should be the household head or another adult whose family migrate to Chiang Mai from his/her hometown. And the interviewee should know well the situation of his/her family and households in hometown.

The survey was conducted in January and February in 2011. It spends the New Year and Chinese New Year. The holiday makes the interview more precise and comprehensive because of the family reunion during the festival.

The information was got and checked by interviewer. The time for each interview was set up to last between forty minutes to one hour. Besides questions on the survey, interviewer carried on the investigation in the dwelling to get the general basic information of households, such as the size and facilities of dwelling. Emotion answers were checked by chatting with the interviewee. The missing variables were also marked.