CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Design

This is a descriptive cross-sectional study, which included two parts. The first section applies nutritional model to explain and evaluate the current nutritional status in China. It will stress on the indicators of preschool children growth and household food intake. A multiple factors analysis will be used in the second part, which will focus on the relationship between nutritional evaluation and household food consumption behavior, then combine relative nutrition policies to develop some policy implication under this study.

3.2 Indicators and Determinations for Description of Nutritional Status

3.2.1 Indicator for Measuring Growth of Preschool Children – Z-score

The best overall indicator of socioeconomic development is the growth performance of preschool children for populations that have not yet met their basic need. This is because the genetic potential of mean preschool children growth is so similar across all genetic populations, including the Chinese, that major differences of growth between populations is always due to impoverished environmental influences. Thus, preschool children growth is a better indicator than infant mortality rate at the developing country where infant mortality rate is poorly measured (J. P. Habicht, Cornell University, 1992).

Growth is an important indicator of health and nutritional status in young children, especially in preschool children, and is the best summary indicator of children well-being. It is also a sensitive indicator of health and nutritional status for general population, because children are the most vulnerable group to health and nutritional threats. Changes in growth reflect overall social and economic development, which affect nutrition through changes in food consumption and in health.

To evaluate the nutritional status of people, growth of preschool children calculation is adopted frequently. Usually, Z-score included 3 indicators, WAZ (Z-score of weight for age), HAZ (Z-score of height for age) and WHZ (Z-score of weight for height). Because the Z-score is age standardized, it allows us to compare Z-score across age groups.

a) WAZ (Z-score of weight for age), if the figure of measurement was less than the standard weight for relative age (standard mean minus 2 standard deviation), also we can say, WAZ less than minus 2 standard deviation to the growth performance of preschool children is underweight. This could represent the acute malnutritional status for preschool children. It was integrated norm to measure the nutritional status of preschool children by UNICEF recommended also.

b) HAZ (Z-score of height for age), if the measurement value was less than the standard height for relative age (standard mean minus 2 standard deviation), or HAZ less than minus two standard deviation, it means stunted to the growth performance of preschool children. It represented the chronic malnutrition of children's nutritional status in a long-term effect.

c) WHZ (Z-score of weight for height), likewise, if WHZ value was less than minus two, that means the weight of measurement is lower than the weight recommended of two standard deviation. It represents that the growth performance of preschool child is wasted. On the contrary, if WHZ value was more than two, that means obesity for growth performance of preschool children.

They were primary indicators in representing the growth performance of preschool children, it was recommended by WHO also. Certainly, another crucial reason influence on the growth performance of preschool children was from genetic side. As is well know, it has obvious difference in the growth performance of people from genetic side between Asian and European. In usage of Z-score, researcher usually adjust the judgement of Z-score in order to avoid the bias from genetic side. For instance, the standard of underweight for preschool children, we could use -2 Z-score for Asian countries and choose –2.5 or –3 Z-score for European countries in measuring the growth performance of preschool children. The specific values of Z-score were listed in Appendix Table A.1 and Table A.2, the calculation method would be mentioned in detailed at section 3.4.1.

3.2.2 Indicator for Nutrients Calculation

In nutriment description, energy, protein and fat were three principal nutrients to maintain the normal physiological function to people's body. Usually, as sensitive indicators for nutritional level, the absolute value of them will be used in responding the nutritional status from dietary intake. According to the standard of nutriments intake which was recommended by MOH of China, the standard of energy intake was 2,400 kcal / per capita per day, and it is called RDA

(Recommended Dietary Allowance) also. Likewise, the standard of protein intake was 70 g / per capita per day and the standard of fat intake was 65 g / per capita per day. They were used to interpret the nutritional quality of the diet. Another critical indicator was percentage of energy from fat. Since Chronic diseases such as diabetes and heart disease increase as the energy from fat rises above 20% of total energy intake, it would be better to keep the proportion less than 25%. WHO recommended dietary fat intake should not higher than 30% of the total energy intake.

3.2.3 Indicator for Integrated Nutritional Evaluation

For assessing the integrated nutritional dietary, the diversification and equilibrium of food intake are regarded as crucial prerequisite. As a very important indicator about integrated nutritional evaluation, the method of DDP (Desirable Dietary Pattern) was used in this study. It was advanced by specialist committee of Asia-Pacific Food and Nutrition Network in 1988.

According to Table 3.3, all food was divided into 8 categories. We could get the percentage of energy share for each category from the calculation of food source of energy. Energy share (%) in Table 3.2 was suitable value of energy providing for each food category, which was figured out by nutritionist in combining food consumption pattern and dietary recommended value. At the same time, based on the different content of nutrients, a rating was given to each category food from 0 to 2.5. Actually, it was a weight value, which represented the importance for each category food in nutritional dietary. There were some principles to be considered in determining on the value of rating, including tasty of the food, energy density of the food, quantity and quality of nutrient content of the food and other empirical judgements. The maximum score of DDP was 100, more close to 100 means the nutritional status was better, if there were some cases was less than 70, then that means the nutritional evaluation and food pattern of those cases were poor. In addition, if the score of each category would be more than its maximum limit of score (in Table 3.2), to get the maximum limit of score only and to come into being deviation score of DDP for the item. Since if the proportion of one specific category food was too large in integrated food dietary pattern, it would cause losing of the tasty for the food, and to be difficult to digest, up to waste nutritional components. Hence, the deviation score of DDP is usually adopted to describe the integrated dietary pattern of cohort from negative side. It means people should improve their food consumption if the deviation score of DDP were too high.

DDP score, which from positive side of nutrition and deviation score of DDP, which from negative side were involved in this part. They used to measure the integrated nutritional status, especially in malnutrition which included under-intake and over-intake for the specific food category. The calculation method will be introduced in detailed at section 3.4.2.

3.3 Conceptual Framework

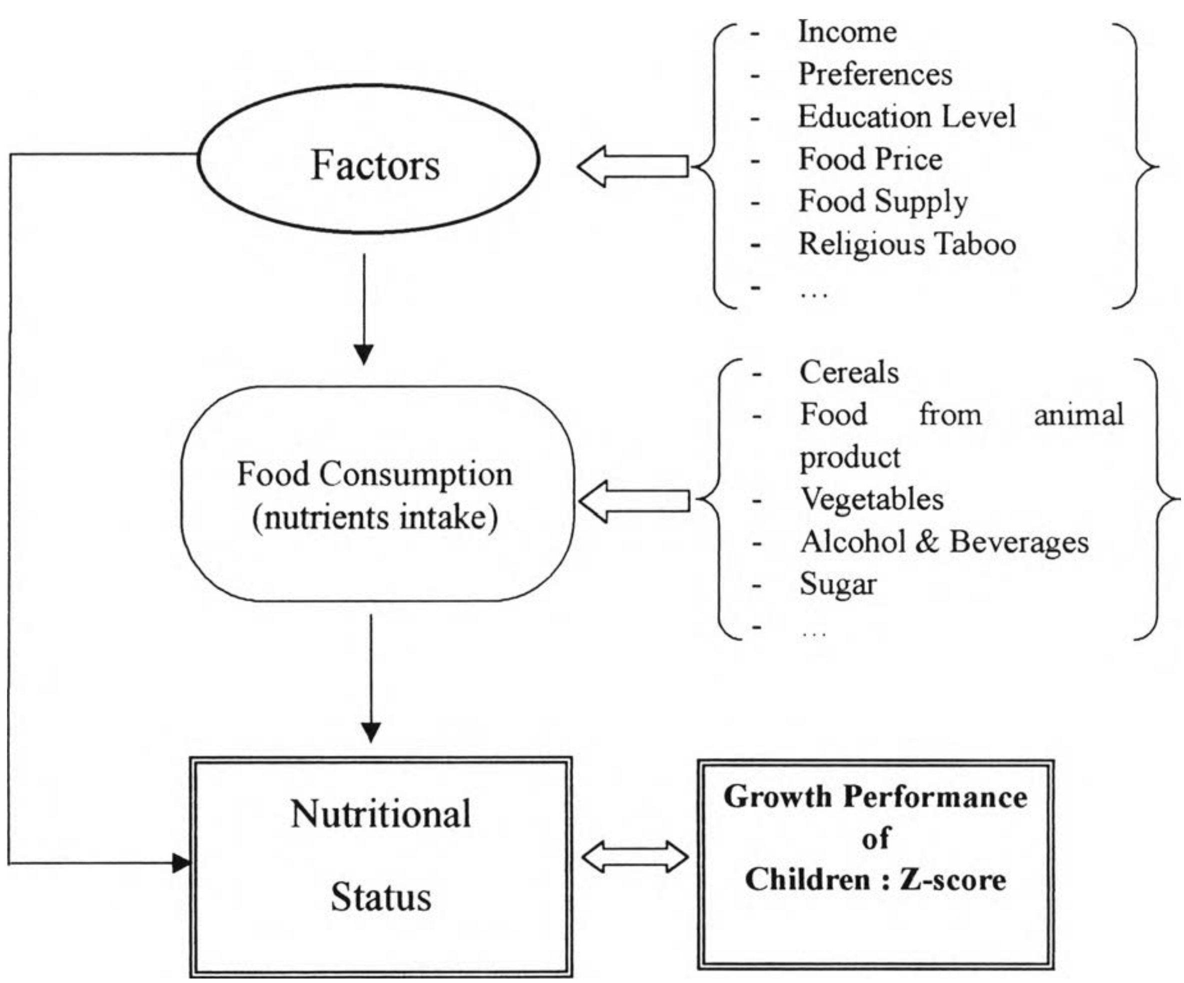
As a crucial indicator to measure the healthy situation and social welfare of people, nutritional status of cohort usually was mentioned in a lot of socioeconomic study. In nutrition, a lot of indicators could be used to point out the nutritional status of people, such as major nutrients intake level, the growth performance of children and adults, integrated dietary evaluation and so on. Among the rest, major nutrients intake level and integrated dietary evaluation would show us the nutritional status of cohort directly. We could figure out the result of nutrient intake and integrated dietary evaluation according to food classification with their nutrient content. Certainly, the source of the food calculated was people's food consumption.

In general, there are many factors to determine people's food consumption. For instance, household income, preferences of consumer in food consumption, accessibility of the food consumption, educational level of people, food price, etc. Herein, food price not only means the own price of specific food, but also including the price for complement and substitute. In some special region, the taboo of religion was determinant in some specific foods consumption. For example, in inhabiting region of Muslim, pork consumption was usually close to zero. Likewise, for traditional Tibetan, they never eat fish. Therefore, it was necessary to use multiple factor analysis to reveal the relationship between nutritional status and food consumption.

Even as we mentioned above, major nutrients intake level and integrated dietary evaluation would show us the nutritional status of cohort directly. On the other hand, as indirect indicators, the growth of performance for children and adults could open out the nutritional status of cohort also. The growth of performance of preschool children was more sensitive to mirror the change of nutritional status of cohort. Of course, it was required to take correlation analysis to show the relationship among the indicators if we would combine the results from different sides.

Two sections would be involved in this study. One is analysis of nutritional status and nutritional evaluation. Z-score for growth performance of preschool children, the intake of energy, protein and fat by income level, the food source of energy, protein and fat by income groups, and integrated nutritional evaluation – DDP and DDP deviation will be adopted in this part. By descriptive nutritional analysis, we would combine the result in each category analysis to identify the current nutritional status and the food dietary pattern. Another section is multiple factor regression analysis between nutritional status and household food consumption as well as other socioeconomic factors. A set of models will be established to illustrate the relationship between nutritional status and some related factors in the context of the following scheme (Figure 3.1). DDP deviation score for major foods will be dependent variables which will be calculated in first part by urban areas and rural areas. At the same time, the sample linear correlation analysis between Z-score and DDP will be developed in this study in order to show the relationship among different method in measuring the nutritional status of cohort also.

Figure 3.1 Conceptual Framework



3.4 Description for Data Analysis

3.4.1 Calculation for Growth of Preschool Children

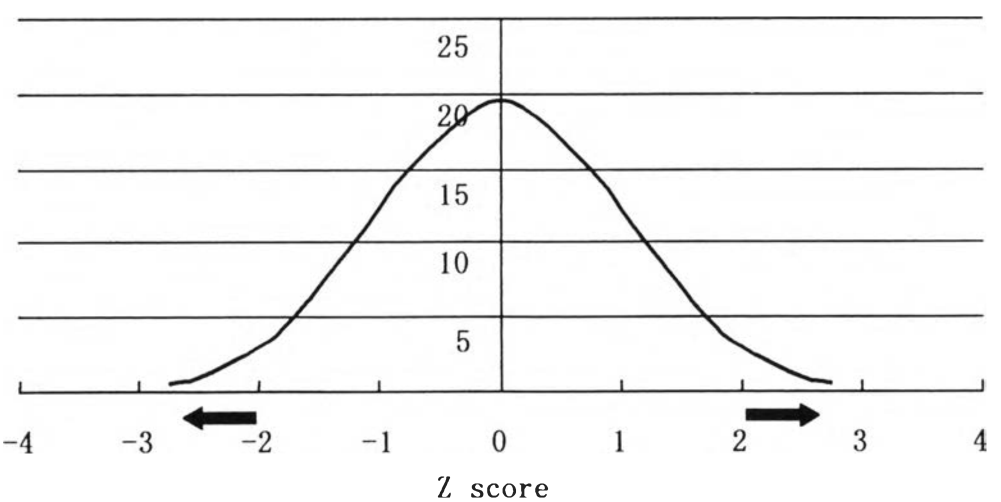
In this study, Z-score will be used to measure the growth of preschool children, which included three indicators, WAZ (Z-score of weight for age), HAZ (Z-score of height for age) and WHZ (Z-score of weight for height). Also, it was recommended by WHO. This method was quite popular in application for nutritional evaluation research.

Figure 3.2 Definition of Z-score

$$Z = \frac{\text{Measurement} - \text{reference standard (WHO)}}{\text{Standard deviation of reference standard}}$$

Weight for age	WAZ < -2SD:	***Underweight***
Height for age	HAZ < -2SD:	***Stunted***
Weight for height	WHZ < -2SD:	***Wasted***
	> 2SD:	***Obesity***

WHO reference



3.4.2 Nutritional Calculation

According to the different family income group to calculate the nutritional data, since the food pattern for each income group household should be different. To use percentile were shown in Table 3.1 to separate whole observations into five equal income groups by household income, each group had twenty percent.

Table 3.1 Criteria for Separating Household Income Group

Grouping by Percentile	Percentile	Household Income (unit: Yuan*)	
		Urban	Rural
Group 1: < 20%	20th	10,650	7,269
Group 2: 20% - 40%	40th	14,232	9,950
Group 3: 40% - 60%	60th	18,002	12,977
Group 4: 60% - 80%	80th	23,502	17,451
Group 5: > 80%			

* Yuan: the unit of Chinese currency. Exchange rate in 1998: 1 US$ = 8.3 Yuan.

In nutrition, energy, protein, and fat are the most important nutrients for people. Both nutritionist and public wanted to know how many people had already gotten them and where them came from, what the food source they were. Usually, using them as the key indicators to measure the nutritional level for the research data set unit, which in this study is household. In additional, the Desirable Dietary Pattern (DDP) will be used as integrated nutritional evaluation in this study.

- Indicators for nutritional dietary:

  - The intake of Energy, Protein and Fat by income level
  - The food source of Energy, Protein and Fat by income level
  - Integrated nutritional evaluation - DDP

1) Definition of DDP

Table 3.2 Desirable Dietary Pattern

	Energy share % (recommended)	Rating	Score	Maximum limit of Score
1. Cereals & Tubers	60	0.5	30.0	40
2. Meat & seafood	14	2.5	35.0	40
3. Added Fats & Oils	9	1.0	9.0	10
4. Legumes & Its Products	5	2.5	12.5	18
5. Sugar	5	0.5	2.5	5
6. Nuts & Oilseeds	2	0.5	1.0	5
7. Fruit & Vegetables	5	2.0	10.0	15
8. Alcohol & Beverages	0	0.0	0.0	0
Total	**100**		**100**	

2) Methodology for the calculation of DDP score

DDP Score = Energy share × Rating

DDP Score$_{max}$ = Maximum limit of Score.

Deviation score of DDP = DDP Score – DDP Score$_{max}$

For example, if the energy share of cereals equal 85 percent of total energy, the score of cereals is 85 × 0.5 = 42.5 > 40, so the score equal 40 only ,and the deviation score of DDP equal 42.5 – 40 = 2.5.

Normally, the DDP score was positive and have no unit, the more close to 100 means the nutritional evaluation was better. If the DDP score of some cases was less than 70, that means the nutritional evaluation for those cases were very poor.

3) Method about calculation of Deviation DDP Score:

The deviation score of DDP was used to represent the nutritional status from negative side. If people consumed large numbers of specific category food, the DDP score of the side food would be more than its maximum limited score. Then, it generated the deviation score of DDP from **over-intake** of the specific food category. On the contrary, if the quantity for a particular food people consumed was too little to catch the recommended score of DDP, likewise, it brought the deviation score of DDP, but from **under-intake** of the specific food category.

For instance, if percentage of the energy share for cereal equal 45, for meat & seafood is 20, then, the deviation DDP score of cereal was,

$$45 \times 0.5 = 22.5 < 30,$$

$$22.5 - 30 = -7.5.$$

The deviation DDP score of meat & seafood was,

$$20 \times 2.5 = 50 > 40,$$

$$50 - 40 = 10.$$

If value of deviation score was negative, it means that deviation came from under-intake side. On the contrary, positive deviation score indicated the quantity on food consumption of specific was more than the standard recommended.

Just as we mentioned above, the deviation score of DDP were from over-intake side and under-intake side. In nutritional evaluation, some of DDP deviation score from both over-intake and under-intake belong to normal aberration, such as deviation score of DDP from over-intake of cereals or fruit & vegetables, DDP deviation score from under-intake of meat & seafood. In nutrition research, there was no distinct evidence to show it has doubtless negative effect to people's health if people eat cereals, fruit or vegetables a little bit more. That means, it was harmless to people's body even the quantity of cereal food or fruit & vegetables people consumed was higher than standard recommended. In this study, we called them **normal deviation**. On the other hand, some of deviation score of DDP had negative influence to people's health, like deviation score of DDP from over-intake of meat & seafood, under-intake from cereal. For this case, we called them **risky deviation**. In order to identify the classification of DDP Deviation for each food category, we showed them in Table 3.3.

Table 3.3 Classification of DDP Deviation by Food Category

Food Category	Deviation Score	Deviation Classification
1. Cereals & Tubers	$\geqslant 0$	Normal
	< 0	Risky
2. Meat & Seafood	$\leqslant 0$	Normal
	> 0	Risky
3. Added & Oils	$\leqslant 0$	Normal
	> 0	Risky
4. Legumes &	$\geqslant 0$	Normal
its Products	< 0	Risky
5. Sugar	$\leqslant 0$	Normal
	> 0	Risky
6. Nuts & Oilseeds	$\geqslant 0$	Normal
	< 0	Risky
7. Fruit & Vegetables	$\geqslant 0$	Normal
	< 0	Risky

3.4.3 Multiple Factors Analysis

1) To find the correlation between DDP score and Z-score.

DDP ~ (HAZ, WAZ, WHZ)

Where,

HAZ	Z_score of height for age
WAZ	Z_score of weight for age
WHZ	Z_score of weight for height

Z-score for growth performance of preschool children and DDP were adopted in assessing the nutritional status this study at the same time. Actually, DDP was the indicator that could point out integrate dietary level directly in a simple way, and it was easy to understand its result. For Z-score, it was able to measure the growth performance to preschool children. Since children, especially for preschool children was the most vulnerable group to unhealthy environment and malnutritional threats, the growth performance of preschool children was a sensitive indicator of health and nutritional status for general population.

Now that, both DDP and Z-score would be used to in measuring the cohort's nutritional status, so that we should answer the question about whether there is any correlation between the two indicators or not. Not only in qualitative but also in quantitative. It means, if they have some correlation, then is it positive or negative and how much?

2) Multivariable regression analysis

In order to point out the relationship between nutritional evaluation and food consumption, a set of multiple linear regression models will be estimated by the method of OLS (Ordinary Least Squares).

$$D = f(INC, EDU, SIZE, FP, \ldots)$$

Where,

D DDP deviation score for major foods, including cereals, meat & seafood and fruit & vegetables

INC Household income

EDU The education level for both household head and the spouse,

SIZE Household size,

FP The price of basic foods.

Cereals, meat & seafood and fruit & vegetables were important foods source for the intake of major nutrients. At the same time, in integrated dietary pattern evaluation, meat & seafood and fruit & vegetables had higher rating, 2.5 for meat & seafood and 2.0 for fruit & vegetables, for cereals had top recommended value of percentage of energy share (60%). It showed us that, the change of those three categories food we mentioned above were more sensitive indicators to influence on people's food consumption whatever on positive side and negative side. Therefore, as dependent variables, we put them into the models estimated individually by urban and rural.

From empirical evidence, the household income is very significant constraint in food consumption, particularly in low-income group. Along with the increment of household income, people would have enough money to satisfy their need in food consumption besides of the basic food security. Certainly, it would be useful to improve their integrated nutritional status if they would take their food consumption in right way. in general, the coefficient of household income was positive number. It means that, with the increment of household income, people will consume more on their food consumption.

The education level of household head and the spouse present that how many nutritional knowledge and nutritional information had been already gotten in household. At the same time, it was able to show us, is it easy or difficult to get something more about nutritional knowledge and to comprehend them in the future. In addition, in China, the household head for most of families is man, especially in rural area. However, usually purchasing food is a task for woman, so that this is the reason to put education level for the spouse into the models that would be estimated. If the household head and the spouse could get high educational level, it should be useful to guide people's food consumption. Since we adopted DDP deviation as our dependent variable, both coefficients of the education to household head and the spouse should be negative number.

The household size is more frequently to be put in nutritional model. Normally, if there were so many members in one family, particularly for the traditional family in rural area of China, they would share the limited household income. Of course, they need much more food to fit their basic need in food consumption. For the same income group, the nutritional status would be cut down inevitably to the family which have more family members. Therefore, the coefficient of household size should be negative number in our model.

We would concern with the food prices for some basic foods including rice, , pork, eggs, vegetables, oil and so on. It is very sensitive indicator in people's food consumption combining their income. As the price of particular food changes, not only will consumption of that food change but consumption of 'complements' and 'substitutes' will also change. A complement is a food directly linked in consumption with the food studied; a substitute is a replacement. The prices for basic foods are significant factors in most of cases about nutrition studies. As the own-price of specific food and the prices for complement foods, the coefficients of them should be negative figure. For substitute foods, the coefficients should represent positive of numbers. In addition, all prices of food in this study were 'implied price'. It means that, the price of specific food was figured out by using the expenditure of the specific food divided by its quantity, those data were collected by household survey.

3) Model specification

In fact, there are not any perfect criteria about model selection in nutrition and food consumption study. We could assume the demand function was exponential relationship and to transform exponential regression model to log-linear model when people wanted to measure the elasticity.

Deviation score of DDP for major foods were figured out from the food consumption that means the quantity of relative foods. We considered that the relationships between household income, food price and food consumption to be nonlinear, so that we just simulated demand function to use a logarithmic transformation of our data. In addition, constant elasticity models in which the dependent variable and the key independent variable are both transformed into logarithms are frequently used in food demand studies to estimate the price and income elasticity.

In this model (often termed a semi-log model), each estimated slope coefficient directly measures the expected percentage change in the dependent variable due to a 1% increase in each of the explanatory variables, holding constant all of the other explanatory variables. If an explanatory variable is not expressed as a logarithm, just like household size, then the coefficient no it measures the expected percentage change in the dependent variable due to a unit change in the explanatory variable, holding constant the other explanatory variables in the model. Other important statistical issues are handled with the cross-sectional modeling used in this study.

A set of semi-log models was used to measure the relationship between dependent variable and each explanatory variable by urban and rural data set. Here, the model just be described in following,

Figure 3.3 The Estimated Multivariate Regression Model

$$\text{Log(RISKY)} = \beta_0 + \beta_1\text{Log(INC)} + \beta_2\text{H\_SIZE} + \beta_3\text{H\_EDU} + \beta_4\text{S\_EDU} + \beta_5\text{Log(P\_RICE)} + \beta_6\text{Log(P\_PORK)} + \beta_7\text{Log(P\_VEG)} + \beta_8\text{Log(P\_EGG)} + \beta_9\text{Log(P\_OIL)} + \mu \quad \text{.........(1)}$$

$$\text{Log(NORMAL)} = \gamma_0 + \gamma_1\text{Log(INC)} + \gamma_2\text{H\_SIZE} + \gamma_3\text{H\_EDU} + \gamma_4\text{S\_EDU} + \gamma_5\text{Log(P\_RICE)} + \gamma_6\text{Log(P\_PORK)} + \gamma_7\text{Log(P\_VEG)} + \gamma_8\text{Log(P\_EGG)} + \gamma_9\text{Log(P\_OIL)} + \upsilon \quad \text{.........(2)}$$

Where:

RISKY	:	Risky Deviation score of DDP for major foods to each household
NORMAL	:	Harmless deviation score of DDP for major foods to each household
INC	:	Household income (Yuan/capita/year)
H_SIZE	:	Household size in number of persons
H_EDU	:	Education level for the household head in number of years
S_EDU	:	Education level for the spouse in number of years
P_RICE	:	Price of rice (Yuan/kg)
P_PORK	:	Price of pork (Yuan/kg)
P_VEG	:	Price of vegetables (Yuan/kg)
P_EGG	:	Price of eggs (Yuan/kg)

P_OIL : Price of oil (Yuan/kg)

Log : Logarithm

$\beta_0 \ldots \beta_9$: Coefficients to be estimated in model 1

$\gamma_0 \ldots \gamma_9$: Coefficients to be estimated in model 2

μ , υ : The error terms

Table 3.4 Expected Sign of Each Explanatory Variable

Explanatory Variables	Cereals		Meat & Seafood		Fruit & Vegetable	
	Normal	Risky	Normal	Risky	Normal	Risky
INC	+	+	+	+	+	+
H_SIZE	-	-	-	-	-	-
H_EDU	-	-	-	-	-	-
S_EDU	-	-	-	-	-	-
P_RICE	-	-	+	+	+	+
P_PORK	+	+	-	-	+	+
P_VEG	+	+	+	+	-	-
P_EGG	+	+	-	-	+	+
P_OIL	+	+	+	+	+	+

Often the degree of sensitivity of a dependent variable to independent variables is represented by a measure called an 'elasticity', the percentage change in the dependent variable resulting from a 1 percent change in the explanatory variable. In order to discuss the result more concisely, we would borrow 'elasticity' as a general word to narrate each logarithmic explanatory variable within section 4.3.2 whatever for household income or food price. In addition, given common economic assumption about price responses, an increase in the price of a food tends to drive consumption away from that food (and its complements) and toward its substitutes.

4) Chow test

Finally, using Chow test to test whether there is any difference or not between the observations had risky deviation score and the observations had normal deviation score in different model. Given the assumptions of Chow test, it can be shown that,

Figure 3.4 Formula for Chow Test

$$F = \frac{(RSS_{restricted} - (RSS_{sub1} + RSS_{sub2}))/K}{(RSS_{sub1} + RSS_{sbu2})/(N_1 + N_2 - 2K)}$$

Where,

$RSS_{restricted}$	the residual sum of squares restricted,
RSS_{sub1}	the residual sum of squares of sub-model$_1$,
RSS_{sub2}	the residual sum of squares of sub-model$_2$,
N_1	the sample size of sub-model$_1$,
N_2	the sample size of sub-model$_2$,
K	the number of parameters estimated.

Follows the F distribution with df = $(K, N_1+N_2 - 2K)$ at the chosen level of α to reject or accept the hypothesis of the regressions of sub-model. If the F from Figure 3.5 exceeds the critical F value at the chosen level of α, reject the hypothesis that the sub-model1 and sub-model2 are the same, that is, reject the hypothesis of structural stability. Alternatively, if the p value of the F obtained from Figure 3.5 is low, then, reject the null hypothesis of structural stability.

3.5 Data Collection

There are two parts of samples in this study:

a). Household food consumption data came from SSB. Stratified random sampling method was adopted in sample survey. The household information, food consumption and relative food price were involved in this data set. Using registration method to finish data collection in 1998.

b). Growth of children data came from CFNSS. To adopt multistage cluster sampling method in sampling survey. The variables included sex, age, height and weight for preschool children as well as education, occupation, income, safety drinking and adult chronic disease for presents and other household members. To use questionnaire survey in data collection at each surveillance point in 1998.

Both of them involved 26 provinces, 40 counties (urban 14, rural 26, including 17 rich counties and 9 poor counties in rural area), more than 16,000 children, number of urban household was 953 and that in rural was 2010.