CHAPTER 6

SURVEY RESULT: A NUMERICAL EXAMPLE

Chapter 5 presented the method to select the optimal premium for VHI scheme in Haiphong, which is used as one of the means to increase the membership of this program. In this chapter, by using data collected from a small pilot survey in Haiphong, the author wants to demonstrate the research methodology on certain numerical examples following the steps of the study described in the chapter 5, adding some more analysis and explanation about the real situation implemented in the study.

6.1 **Analyze the Historical Data and Behavior of Population to Select the Item for Survey's Question**

Basically, the analysis of historical data in the step 1 of chapter 5 and here are the same, so there is no need to repeat it again. It shows that it is better if we choose a premium which is lower by 30-35% compared to the current premium of 15,000 dong and set it into survey questions. Note that, in fact, because of hyperinflation in the previous time, the behavior of Vietnamese now is counting, thinking and using currency units with 500, 1,000, 5,000, 10,000 dong...So, we choose a premium of 10,000 dong for the survey question as the most appropriate selection. In case of mentioning about their income, the Vietnamese always say "my income per month is about 300,000; 500,000; 1,000,000...". No body says his or her income is 450,000; 370,000 or 550,000... It is an other point which should be applied in asking the household income in the survey.

6.2 **Carry out the Survey to Collect Data for Running Logit Model**

The survey question form was described in detail at chapter 5. Here, the author emphasizes to present some more task of operating the survey questions.

6.2.1 **The Survey's Data Collection**

The survey was carried out in December 1995 at different places in Haiphong, including 3 urban district: Ngo Quyen, Le Chan and Hong Bang and 3 rural districts : Thuy Nguyen, An Lao and An Hai, by asking people who went to schools to pick up their children, or asking patients who visited commune health stations, clinics and hospitals.

6.2.2 **Sample Size**

The purpose of the survey was to collect data to roughly depict steps of the study presented in chapter 5. Since the time is limited, so the sample size of survey was rounded to 200. The degree of accuracy required for this survey is about 0.07 according to the formula (5.7) shown in section 5.5.2, with sample size is 200.

$$n = \frac{Z^2_{\alpha} \cdot p \cdot q}{d^2} \qquad (5.7)$$

$$n = \frac{(1.96)^2 \cdot (0.5).(0.5)}{(0.07)^2} = 196, \text{ nearly} = 200$$

For random collection, the resident location of interviewees were 98 in urban and 102 in rural areas.

6.2.3 **Summary of Survey Data**

Firstly, summary survey data were set into a summary table, then keyed into the computer using lotus software which will be used for running logit model with TSP7 software, in order to find the logarithm odds ratio-regression Y.

$$Y = \beta_0 + \beta_1 INC + \beta_2 AGE + \beta_3 EDU + \beta_4 FAMS + \beta_5 LOCA \qquad (5.8)$$

where

INC is income variable of household.

AGE is age variable of householder.

EDU is education level variable of householder counted by the years studying in school and University.

FAMS is family size variable, included householder, husband or wife and his or her dependent.

LOCA is resident location binary variable of household, with (1) is urban, (0) is rural.

The results of summary table are presented in Appendix 2.

6.3. **Analyze Survey's Data by Running Logit Model for Finding the Logarithm Odd Ratio-Regression Y**

$$\text{Ln}\frac{Pi}{1-Pi} = Y = \beta_0 + \beta_1 INC + \beta_2 AGE + \beta_3 EDU + \beta_4 FAMS + \beta_5 LOCA \quad (5.8)$$

where Y is the logarithm odd ratio-regression.

6.3.1 **Run Logit Model**

Use Software TSP7 to read the survey data file assembled in Lotus 123r24 software, we can get the results as follows (Table 6.1):

Table 6.1: Results of Survey Data's Estimated from Logit Model

VARIABLE	COEFFICIENT	STD.ERROR	T-STAT	2-TAIL SIG
Constant	-25.173	5.053	-4.98	0.000
INC	5.2E-06	2.34E-06	2.22	0.027
AGE	0.362	0.073	4.96	0.000
EDU	0.782	0.182	4.30	0.000
FAMS	-1.017	0.484	-2.10	0.037
LOCA	-0.923	0.761	-1.21	0.227

Log likelihood = -27.304

Y is dependent variable.

The detail of these results of survey data estimated from logit model are presented in *Appendix 3*.

6.3.2 **Test for Significance of Each Variable of the Regression Y**

Test the hypothesis that each coefficient of Y is different from zero, with :

$H_0 : \beta_i = 0$
$H_1 : \beta_i \neq 0$ $\qquad i = 1,2...5$

The meaning of this test is if the null hypothesis H_0 is true, the corresponding independent variable is not related to regression Y, and its value is useless; in the opposite case when alternative H_1 is true, we can conclude that there is a relationship between that variable and regression Y.

For two-signed test of significance, the critical value from the *t table* = ± 1.96 for $\alpha = 0.05$ and 200 degrees of freedom.

The results of logit analysis computed by TSP software in the section 6.3.1 above show us that :

i) Among 5 independent coefficients, 4 of them, β_1, β_2, β_3, β_4 corresponding with income, age, education, family size respectively, have absolute computed test values (t-statistic) which are bigger than 1.96. So for these, we reject H_0 for these variables and conclude that these variables have statistical meaningful relationship.

ii) In case of coefficient β_5, because its t-statistic = -1.212 > -1.96, we have to erase β_5 and conclude that there is no statistical meaningful relationship between the variable LOCA and regression Y.

From those results, we can rewrite the regression equation
$Y = \beta_0 + \beta_1 INC + \beta_2 AGE + \beta_3 EDU + \beta_4 FAMS + \beta_5 LOCA$ as follows :

$$Y = \underset{(-4.98)}{-25.173} + \underset{(2.22)}{0.0000052 INC} + \underset{(4.97)}{0.362 AGE} + \underset{(4.30)}{0.782 EDU}$$

$$- \underset{(-2.10)}{1.017 FAMS} - \underset{(-1.21)}{0.923 LOCA}$$

where the t-values for the estimates are presented in parentheses.

6.4 The Hypothetical Situation for Membership of VHI Program when They Sell VHI Card at Lower Premium (10,000 dong)

6.4.1 Estimate the Value of Regression Y and Probability Pi

As the analysis in the point (i) of section 5.5.4 of chapter 5, depending on the purpose of analysis, a lot of values of Y can be derived from different value suggested for independent variables INC, AGE, EDU, FAMS, base on the secondary or predicted data about the real situation related the above variables.

After estimating the value of Y, anti-$\log_e$ for two size of equation (5.8)

$$\text{Ln}\ \frac{P_i}{1- P_i} = Y \qquad (5.8)$$

we can be rewritten the equation (5.8) as equation (5.9)

$$\frac{P_i}{1- P_i} = \text{anti-Log}_e\, Y \qquad (5.9)$$

From equation (5.9) we can get the hypothetical value of P_i which presents the probability of the population who are willing and able to buy VHI card with lower premium suggested (10,000 dong).

If we suggested a lot of various situations with different values of variables, firstly we get a lot values of Y, then after that we will get a lot of corresponding values of probability Pi. Total number of VHI program membership in that case can be predicted by summing values of each probability Pi multiplied by each number of corresponding classified population groups.

Now, we simulate some cases from them.

Suppose we get the data of evaluation related to income of sub-groups among the target population of VHI program in Haiphong, with corresponding age, education level and family size, then when those data are assembled

into regression Y, the values of corresponding probability (Pi) can be calculated.

Table 6.2: Simulated Different Situations of Variables, Y, Pi among Target Population of VHI Program in Haiphong

Group	Low income		Middle income		High income	
Sub-group / Item	1	2	3	4	5	6
1.Household income (*thousand dong*)	600	600	900	900	1,200	1,200
2.Age of householder (*year*)	45	50	45	45	40	40
3.Education level (*year*)	9	9	10	10	12	12
4.Family size (*) (*person*)	4	5	5	4	5	4
6.Value of Y	0.061	0.136	0.231	0.639	0.860	2.377
7.Value of Pi	0.058	0.120	0.188	0.390	0.413	0.704

(*) Includes householder, spouse and their dependents.

Simulative analysis :

With above data and results of Pi,

(1) Among low income levels, even index of family size is bigger, but sub-group (2) gets value of Pi = 0.12 > 0.058 from sub-group 1. The affected factor may come from age variables, which the manager of VHI program need to pay attention to.

(2) At middle income levels, with income higher by 1.5 times and other indices nearly the same, sub-group (3) has a value of Pi = 0.19, 1.5 times an 3 times higher than sub-group (1) and (2), while sub-group (4) whose family size is lower than sub-group (3) 1 person, get Pi = 0.39 which is double.

(3) At high income levels, with indices of income of 1,200,000 dong and education level 12, those factors strongly influence the value of Pi, which are 0.413 and 0.70 for sub-group 5 and 6; it is very high compared to the others.

(4) For implementation, if we know the number of population for each sub-group, we can be better able to predict the number of people who should buy VHI card, by using each probability Pi multiplied by each corresponding population sub-group.

6.4.2 **Choose Value of Probability Pi**

The purpose of the study in chapter 6 is to demonstrate the research methodology described in chapter 5, and in the situation of limitation of data and time, we should use the *means* computed from running the logit model as the general representative data to estimate the value of Y, after that to calculate probability Pi, then predict the number of population who are willing and able to buy VHI card in Haiphong.

We used the mean of all of samples computed from the logit model,

where

Constant = -25.173

INC is average family income of respondent = 786,000 dong.

AGE is average age of respondent = 42.860 year old.

EDU is average education level of respondent = 10.665 years.

FAMS is average family size of respondent = 4.100 people.

Using alternative values of INC, AGE, EDU, FAMS into regression Y above, we have :

$$Y = -25.173 + (0.0000052 \times 786{,}000) + (0.362 \times 42.860) + (0.782 \times 10.665) - (1.017 \times 4.100)$$

$$Y = -1.39729$$

After estimating the value of Y, the equation (5.8) can be written as follows :

$$\text{Ln} \frac{P_i}{1 - P_i} = Y = -1.39729$$

Then anti-$\log_e$ two size of the equation (5.8) :

$$\frac{Pi}{1 - Pi} = 0.247266$$

and $Pi = 0.247266 - 0.247266\ Pi$

$1.247266\ Pi = 0.24766$

The final result of probability P_i is $P_i = 0.198246$

6.4.3 **Predict Number of Population Who are Willing and Able to Buy VHI Card (Q_1) with Premium Level 10,000 dong per Card (P_1)**

Following the result of section 6.4.2, Q_1 equals to probability P_i (x) Number of objective insured group

= 0.198246 (x) 1,060,000 = 210,140 people.
rounded to = 210,000 people.

6.5 **Calculate the Coefficient of Price Elasticity η Between Two Point of Demand Curve P_0, Q_0 and P_1, Q_1**

From equation

$$\eta = - \frac{Q_1 - Q_0}{P_1 - P_0} \times \frac{P_1 + P_0}{Q_1 + Q_0} \qquad (5.6)$$

where

P_0 is current premium 15,000 dong;

Q_0 is correspondent current VHI insured, 105,000 people;

P_1 is lower premium suggested for survey, 10,000 dong;

Q1 is the predicted number of people who are willing and able to buy VHI card at premium P_1 calculated in section 6.4.2 = 210,000 person.

We can calculate η :

$$\eta = - \frac{(210,000 - 105,000)}{(10,000 - 15,000)} \times \frac{(10,000 + 15,000)}{(210,000 + 105,000)} = 1.67$$

6.6 **Predict Corresponding Qi when Premium Pi Decreases and Varies from 15,000 dong to the Lower Premium Suggested is 10,000 dong**

From result calculated in section 6.5 above, we can convert the formula (5.6) to the form as follows :

$$1.67 = - \frac{(Qi - 105,000)}{(Qi + 105,000)} \times \frac{(Pi + 15,000)}{(Pi - 15,000)}$$

where

1.67 is the value of coefficient of price elasticity (η)

15,000 dong is current premium (P_0)

105,000 is corresponding current VHI insured people (Q_0)

P_i is the hypothetical value of VHI premium decreases and varies along demand curve from 15,000 dong to 10,000 dong.

Qi is the corresponding predicted number of people who are willing and able to buy VHI card, when Pi decreases and varies along demand curve from 15,000 dong to 10,000 dong.

When we use values of Pi decreasing and varying along the demand curve from 15,000 dong to 10,000 dong as alternatives in this formula, we can predict corresponding Q_i which are used in step 8 for choosing the optimal premium of VHI program in Haiphong. For example :

Premium level **P**i(dong)	Predictive **Q**i (person)
15,000 -	105,000 -
14,500 -	111,117 -
14,000 -	117,832 -
13,500 -	125,236 -
13,000 -	133,443 -
12,500 -	142,588 -
12,000 -	152,844 -
10,000 -	210,140 -

6.7 **Calculate the Costs of VHI Program in Haiphong**

Total cost (TC) of VHI program in Haiphong includes two portions : hospital cost for VHI patients (HC) and VHI administrative cost (AC), which can be computed by 2 formulae described in detail in section 5.5.7 of chapter 5 in the form :

$$HC = \sum_{h=1}^{z} \sum_{j=1}^{m} \sum_{i=1}^{n} \left[\sum_{g=1}^{t} WC_{hijg} + \sum_{r=1}^{u} UC_{hijr} + \sum_{v=1}^{w} SC_{hijv} \right] \qquad (5.9)$$

$$AC = \sum_{x=1}^{b} \left[\sum_{q=1}^{c} AD_{qx} + \sum_{t=1}^{k} MA_{tx} + \sum_{p=1}^{e} SA_{px} \right] \qquad (5.10)$$

For the year 1995, the total hospital cost (HC) (presented as an even number by million dong) calculated by following this formula is 1,225 million, while VHI administrative cost (AC) calculated directly by the VHI office is 105 million (rounded figure).

So, total costs (TC) in 1995 of VHI Program is :

TC = HC + AC = 1,225 + 105 = 1,330 million dong.

6.8 **Choose the Optimal Premium for VHI Program in Haiphong**

Based on the method described in step 6, step 7, and "the Assumption" described in section 5.5.8, we can calculate the indices of membership Q_i; ratio $Q_i/Q_0 = \alpha_i$; total revenue = $Q_i.P_i$; maximum cost $TC_i = TC_0.\ \alpha_i$; surplus or deficit fund... corresponding with many levels of premium when it decreases and varies from 15,000 dong to 10,000 dong, then set those results into the form of Table 6.3 below and compare them with "Criteria of Optimal Premium for VHI program in Haiphong" described in the section 5.1, we can choose the optimal premium for VHI Program in Haiphong in terms of 1995 price.

Table 6.3: Choose the Optimal Premium for VHI Program

Premium **Pi** (1995 price) / Index	15,000 (dong)	12,700 (dong)	**12,672** (dong)	12,600 (dong)	10,000 (dong)
1. # of predicted membership **Qi**	105,000	138,807	139,326	140,676	210,000
2.Total revenue =**Qi.Pi** (million dong)	1,575	1,763	**1,765**	1,773	2,100
3. Ratio **Qi/ Qo = α_i**	1	1.320	1.327	1.340	2
4. Maximum cost **TC_i = TC_o * α_i**	1,330	1,758	**1,765**	1,783	2,660
5. Surplus or deficit (million dong)	245	5	**0**	-10	-560
6. Certified optimal premium of VHI program			*optimal premium*		

Concerning the hypothetical results presented in Table 6.3, a premium level of 12,672 dong satisfies all criteria of the optimal premium for VHI program in Haiphong: Program *revenue equal to maximum costs* predicted, with *surplus equal to 0; increase of 32.7% in membership* of VHI program from 105,000 to 139,326 insured by *decreasing premium* from 15,000 to 12,672 dong with η=1.67 >1, so, we can conclude that *the optimal premium for VHI Program in Haiphong predicted by this study is 12,672 dong at 1995 price.*

We can convert the optimal premium predicted above to 1996 price by the formula as follows :

$$OP_{1996} = OP_{1995} * (1 + r) \qquad (5.11)$$

where

OP_{1996} is optimal premium at 1996 price.

OP_{1995} is optimal premium at 1995 price.

r is estimated inflation rate increasing in 1996 compared to 1995, or estimated increasing percentage of price index in 1996 comparing to 1995.

According to table 2.1 where inflation rate of 1995 was estimated is 10%, that means (1 + r = 1.1)

OP_{1996} = 12,672 x 1.1 = 13,939 dong.