3.1. Introduction

In chapter two we have presented a detailed survey of literature on the studies relating to financing on health care and health insurance scheme. This chapter will deal with the presentation of econometric model of health insurance and the corresponding method of estimation. In chapter one in the context of writing our basic objectives of the study on demand for health insurance we pointed out three important issues that are to be considered for empirical estimation for sample observations in the district of Birbhum. These issues are related to (i) the impact of health insurance on the decision to visit health care centers or the rate of utilization of health cares and services, (ii) the impact of health insurance on out-of-pocket payment for health care expenditure and (iii) demand for health insurance or the decision to purchase health insurance in the district of Birbhum. In order to carry out
the study with our eyes on the basic objectives, we need to consider the basic analytical framework and the method of estimation for each of the above objectives. We need also to specify the actual variables and the practical models that will be used and estimated with the help of data. Therefore, we proceed to organize this chapter as follows.

In section 3.2 the analytical framework of the demand for health insurance has been considered. Under this section, we have three sub-sections. Sub-section 3.2.1 discusses the analytical framework the decision to visit health care centers or the rate of utilization of health cares and services. In sub-section 3.2.2, the analytical model for out-of-pocket payment for health cares and services is considered. Sub-section 3.2.3 deals with the analytical framework for the decision to purchase health insurance. In sub-section 3.2.4, we have presented the method of estimation for the different theoretical and analytical models for the issues relating to the demand for health insurance. The specifications and definitions of the explanatory variables actually used in all the models have been presented in section 3.3. The health consciousness is an important explanatory variable that will affect the decision to visit health care centers or rate of utilization of health cares and services, out-of-pocket payment for health cares and services and also the decision to purchase health insurance. The criterion for measuring the health consciousness has been presented in sub-section 3.3.1. In section 3.4 along with its three sub-sections, we have presented the actual specification of the models on decision to visit health care centers or rate of utilization of health cares and services, out-of-pocket payment for health cares and services and the decision to purchase health insurance. The issue
of endogeneity and selection bias has been discussed in sub-section 3.4.4. In all these models some hypotheses are to be tested empirically. These hypotheses are specified in the section 3.5 in relevant sub-sections. In section 3.6, the relevance of case study in the district of Birbhum is discussed along with the methodology of data collection in sub-section 3.6.1 and the diagnostic check for the sample size in sub-section 3.6.2. Finally the conclusion is presented in section 3.7.

3.2. Analytical Model of Demand for Health Insurance

In order to look into the matter of demand for health insurance in the district of Birbhum we have planned to proceed as follows.

First, let us assess the impact of health insurance on the financial protection of those who came under the umbrella of the health insurance scheme. With this end in view, two aspects are to be considered.

(i) We would like to look into the probability of visiting health care centers and this probability is definitely affected by the decision to purchase health insurance.

(ii) The out-of-pocket expenditure that was shouldered by the persons/households visiting health care centers would differ depending on the fact that the persons/households under consideration are
protected or not protected by health insurance. By the out-of-pocket expenditure we mean the total expenditure on health care and services netted of the amount of reimbursement when the persons/households are protected by health insurance. This amount of reimbursement is naturally zero if the persons/households are not protected by health insurance.

However, from the first part of our objectives, we would be able to examine the determinants of utilization health cares. This will, in turn, enable us to analyze whether participation in the health insurance scheme increases the access to the health care services. In other words, we would like to see whether the umbrella of the health insurance scheme reduces the barriers of the persons/households to receive the health care services.

Given the first objective, our second objective is to estimate the net out-of-pocket payment or expenditure of the persons/households who had to go in for hospital care and services in the district of Birbhum during the period 2002 – 2003.

Second, we plan to analyze and examine what factors are important and significant in the determination of decision to purchase health insurance. In the following subsection we are going to present the econometric models for our objectives as outlined above.
3.2.1. Decision for Visiting Health Care Centers

In order to assess the probability of visiting health care centers by the persons or households in the district of Birbhum we shall use a logit model. The dependent variable in such a situation is the decision of a person to visit health care centers. Decision to visit health care centers is a dichotomous variable and it takes two values – 0 and 1. If the individual decides to go to visit the health care centers, the decision to visit health care centers take unitary value otherwise it takes the value zero in the case of the negative response. Thus as in the case of a typical binary response model, in this case also decision to visit health care centers which is the dependent variable \( Y_i \) can take on only two values 1 and 0.

Let \( P_i \) stand for the (conditional) probability that the persons/households visit the health care centers. In such a case, in terms of a binary response model we try to assess \( P_i \) conditional on certain information set, say \( \Phi_i \), that consists of individual, household and community characteristics. These characteristics are simply the exogenous and predetermined variables that are the determinants of the conditional probability of visiting health care centers by the households. If we specify \( \gamma \), so that it is either 1 or 0, \( P_i \) is then simply the expectation of \( \gamma_i \), conditional on \( \Phi_i \), that is \( E(\gamma_i | X_i, \in \Phi) \), where \( X_i \) stands for individual, household and community characteristics belonging to the information set. Thus, the decision to visit health care centers will depend on information set (\( \Phi_i \)), which may be classified as follows.
Individual and Household Characteristics

Individual and household characteristics refer to a set of most important instrumental or control variables that would affect the decision of persons or households to visit health care centers. We consider the individual characteristics such as age, sex, literacy, health consciousness and frequency of illness as important variables affecting the decision to visit health care centers. There is nothing to deny that the need for health care and services would differ depending on the age of the person. So is true with respect to sex. It is natural that the females are more vulnerable to disease compared to males. It should be pointed out therefore in this context that age, sex and literacy as the individual characteristics will reflect the differences and the extent of the need for the health care use. On the other hand, the health status of the individuals will be revealed by the health consciousness and frequency of illness.

The household characteristics are also important in the determination of the decision to visit health care centers. We include in the spectrum of the household characteristics the variables such as income and region of the households. These variables are instrument in the control of health preferences of the households.

Finally, community characteristics are no less important in affecting the decision to visit health care centers. The religion of the households and the caste the households belong to are incorporated in this spectrum. Religion as the community characteristic of the households serves in the
district of Birbhum as one of the most important control variables in the
decision to visit health care centers. We have observed while collecting the
sample observations for this study that people in the Muslim community
often prefer going to hekims instead of going to registered medical
practitioners. Similarly, most of the households in the scheduled caste and
scheduled tribe do not come to doctors, unless some emergency demands.
They depend mainly on some unscientific mode of treatment rendered by
ojha, gunin etc.

3.2.1A. Analytical Framework for Decision to Visit Health Care Centers

As discussed in the previous section, we may write

\[ \text{Decision to Visit Health Care Centers} = f(\text{Income of the Household, Households Characteristics, Community Characteristics and Random Disturbances}) \]

If we consider a general linear model, where the decision to visit health care
centers \((Y_i)\) is determined by \(X_i\) that includes all the characteristics and the
random disturbances, then we write

\[ Y_i = X'_i \beta + U_i \]
where

\[
P_i = P(Y_i = 1) \Rightarrow \text{The probability that the } i^{th} \text{ household decides to visit health care centers.}
\]

\[
1 - P_i = P(Y_i = 0) \Rightarrow \text{The probability that the } i^{th} \text{ household decides not to visit health care centers.}
\]

The conditional probability that the household decides to visit health care centers, given other things (i.e., \(X_i\)) is equivalent to the conditional expectation that the household will decide to visit health care centers, given \(X_i\). That is,

\[
P_i = \Pr(Y_i = 1|\Phi_i) = E(Y_i|\Phi_i)
\]  \hspace{1cm} (3.2.1)

The objective of the binary response model is to estimate this conditional expectation.

It is clear from the above discussion that the linear regression model carries no sense as a binary response model. Suppose a column vector \(X_i\), which is of \(n \times k\) (\(k\) being the number of explanatory variables and \(n\) being the number of observations) belongs to the information set \(\Phi_i\) including the constant term or the equivalent. Then the linear regression model would specify \(E(Y_i|\Phi_i)\) as a probability and the probabilities must lie between 0 and 1. Since \(E(Y_i|X_i) = X_i^\prime \beta\) and it is equal to the probability that the household or the person will decide to visit health care centers, it should lie between 0 and
1. Several empirical studies simply use OLS to estimate what is called (rather inappropriately) Linear Probability Model (LPM). The basic problem in such linear probability model is that there is no certainty that the probability would lie in between 0 and 1.

Therefore, if \( Y_i \) stands for decision to visit health care centers of the \( i \)-th household and \( X_i \) is a set of explanatory variables (we parenthesize all explanatory variables determining the decision to visit health care centers within \( X_i \)) and if the relationship between \( Y_i \) and \( X_i \) is specified by a linear relationship, we can write

\[
Y_i = X_i' \beta + U_i \tag{3.2.2}
\]

where

\[
Y_i = \begin{cases} 
1, \text{ if the } i \text{-th household decides to visit health care centers} \\
0, \text{ if the } i \text{-th household decides not to visit health care centers}
\end{cases}
\]

We assume that \( E(U_i) = 0 \) for all \( i \) as required in the classical linear regression model. Then, it follows from equation (3.2.2)

\[
E(Y_i|X_i) = 0 \cdot (1 - P_i) + 1 \cdot P_i = P_i = X_i' \beta \tag{3.2.3}
\]
The basic condition we need for equation (3.2.3) is that

\[ 0 < E(Y_i | X_i) < 1 \]

It should be noted that even if \( X' \beta \) happens to lie between 0 and 1 for some \( \beta \) and all observations in a particular sample, it is impossible to constrain \( X' \beta \) to lie in that interval for all possible values of \( X_i \), unless the values that the independent variables can take are limited in some way (for example, they might all be dummy variables). Thus, it seems that the linear probability model is flawed and not appropriate for use.

Therefore, we need to search for other binary response models, which are free of such problems and are easily tractable. In all these models, our principal intention is to transform the function \( F(X) \) such that it has the properties as

\[ F(-\infty) = 0, \quad F(\infty) = 1, \quad (3.2.4) \]

and

\[ f(X) \equiv \frac{\partial F(X)}{\partial X} > 0 \quad (3.2.5) \]

This ensures \( F(X) \) as a monotonically increasing function that maps from the real line to the 0–1 interval. There are many cumulative distribution functions, which are endowed with these properties. Therefore, we would use various specifications for the transformation function and develop the model of the conditional expectation of \( Y_i \) in a number of ways.
The probabilistic model that would assess the decision to visit health care centers can be framed by means of transforming \( X', \beta \) into a probability as follows. Let us use the transformation function \( F \) such that

\[
P_i = P(Y_i = 1) = F(X', \beta)
\]  

(3.2.6)

What we should do is to use a transformation function \( F(X) \) applied to an index function that depends on a set of explanatory variables, which we call the determinants of decision to visit health care centers. We like to estimate the parameters of such a model. An index function is easily tractable and it holds the properties of a regression function, linear or nonlinear. Then we can make a very general specification of a binary response model as

\[
E(Y_i|\Phi_i) = F(h(X_i, \beta)),
\]

(3.2.7)

where \( h(X_i, \beta) \) is our desired index function.

Instead of specifying (3.2.7), we can specify alternatively as

\[
E(Y_i|\Phi_i) = F(X', \beta).
\]

(3.2.8)

What we note in this case is that the index function \( X', \beta \) is linear and \( E(Y_i|\Phi_i) \) is simply a nonlinear transformation of \( X', \beta \). It is clearly visible that
\( X' \beta \) can in principle take any value on the real line. But, it is required that \( F(X' \beta) \) must lie between 0 and 1 for its meaningfulness by property in (3.2.4). As \( F(X' \beta) \) is a nonlinear function, \( E(Y_i|\Phi_i) \) will be affected by the changes in the values of the elements of \( X_i \), say \( X_y \), necessarily in a nonlinear fashion.

Thus if, \( P_i = E(Y_i|\Phi_i) \) is given by equation (3.2.8), differentiating it with respect to \( X_y \) gives

\[
\frac{\partial P_i}{\partial X_y} = \frac{\partial F(X' \beta)}{\partial X_y} = f(X' \beta) \cdot \beta_y. \tag{3.2.9}
\]

For the transformation functions that are almost always employed, \( f(X' \beta) \) achieves a maximum at zero and then falls as \( X' \beta \) gets further from zero. Thus (3.2.9) tells us that the effect on \( P_i \) of a change in one of the independent variables is greatest when \( P_i = 0.5 \) and least when \( P_i \) is close to 0 or 1.

In order to use the binary response models it is required to employ the linear index function \( X' \beta \) along with one or two particular specifications of \( F(\cdot) \). Depending on such specifications our efforts finally result in models, which are called the **Probit model** or the **Logit Model**.

However the probit model is based on the transformation function \( F \) that is simply the cumulative standard normal distribution function and the
logit model is based on the transformation function $F$ that is logistic distribution.

For the probit model we have

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)dx$$

where $x$ is a standard normal variable and $F(x)$ being a cumulative distribution function (c.d.f.), automatically satisfies conditions (3.2.4 and 3.2.5).

The probit model is a nonlinear statistical model (in the parameters). It comes successful in relating the probability $P_i$ to the explanatory factors in such a way that the probability is constrained to remain in the $[0, 1]$ interval. The probit model can also be developed in the case of discrete choice. If we want to develop the probit model for a discrete choice, we need to define the utility index function $I_i$ for the $i^{th}$ individual as

$$I_i = X_i'\beta$$

The value of utility index function and the value of the probability that the $i^{th}$ individual chooses the option where $Y_i = 1$ are directly related.

It should be noted that with this probability model, the probability $P_i$ that $Y_i = 1$ varies between zero and one, since it is the probability that the standard normal variable $X$ is less than or equal to $I_i = X_i'\beta$, which lies
between \(-\infty \) to \(+\infty \). The cumulative distribution function is a probability transformation of \( I_i \) that achieves our objectives of keeping \( P_i \) between zero and one and yields a monotonic relationship between the utility index function \( I_i \) and \( P_i \).

The probit function which represents the choice probability \( P_i \) is

\[
P_i = E(Y_i | \Phi_i) = F(I_i) = F(X_i \beta)
\]  

(3.2.12)

where, \( F(I_i) \) is the cumulative distribution function (c.d.f.) of the standard normal \( N(0,1) \) random variable evaluated at \( I_i \). As mentioned in equation (3.2.10), the cumulative distribution function is thus given by

\[
P_i = F(I_i) = F(x \leq I_i) = \int_{-\infty}^{I_i} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} x^2) dx
\]  

(3.2.13)

In a similar fashion we can develop the logit model. For the logit model, the function \( F \) is the logistic function as follows.

\[
\psi(X) \equiv \{1 + \exp(-X)\}^{-1} = \frac{\exp(X)}{1 + \exp(X)}
\]  

(3.2.14)

Differentiating (3.2.14) with respect to \( X \) gives

\[
\nu(X) = \frac{\exp(X)}{[1 + \exp(X)]^2} = F(X)F(-X)
\]  

(3.2.15)

129
Solving for $P_i$, we have

$$P_i = \frac{\exp(X'_i \beta)}{1 + \exp(X'_i \beta)}$$

$$= \left\{ 1 + \exp(-X'_i \beta) \right\}^{-1}$$

$$= F(X'_i \beta) \quad (3.2.16)$$

Thus, if we choose $F$ to be the logistic distribution, Logit Model can be developed as follows.

$$P_i = P(Y_i = 1) = F(X'_i \beta) = \frac{1}{1 + \exp(-X'_i \beta)}$$

$$= \frac{\exp(X'_i \beta)}{1 + \exp(X'_i \beta)} \quad (3.2.17)$$

$$= \frac{\exp(Y_i)}{1 + \exp(Y_i)} \quad (3.2.18)$$

Equation (3.2.17) represents the cumulative logistic distribution function (Cramer, 1991).
We see that as $Y_i$ varies from $-\infty$ to $+\infty$, $P_i$ varies from 0 to 1 and $P_i$ is non-linearly related to $Y_i$, as shown below.

![Cumulative Density Function](image)

Figure – 3.1. A Cumulative Density Function

From equation (3.2.17) we have

$$P_i = \frac{\exp(X_i^\prime \beta)}{1 + \exp(X_i^\prime \beta)}$$

Hence,

$$1 - P_i = 1 - \frac{\exp(X_i^\prime \beta)}{1 + \exp(X_i^\prime \beta)} = \frac{1}{1 + \exp(X_i^\prime \beta)}$$  \hspace{1cm} (3.2.19)
Therefore, we can write,

\[ \frac{P_i}{1 - P_i} = \exp(X'_i \beta) \]  (3.2.20)

This is called the **Odds Ratio** in favour of decision to visit health care centers. The Odds Ratio is simply the ratio of probability that a household will visit health care centers to the probability that the household will not visit health care centers.

Taking logarithm of the equation (3.2.20), we get,

\[ \ln\left(\frac{P_i}{1 - P_i}\right) = X'_i \beta \]  (3.2.21)

The natural logarithm of equation (3.2.20), that is the Odds Ratio as represented in the above equation is called the Logit and is an exact linear function of \( X \).

### 3.2.2. Out-of-Pocket Payments for Health Care and Services

It is true that all persons or households will not be required to visit health care centers. Once we have developed the model for decision to visit health care centers, it is also our objective to develop a model by which we can estimate the out-of-pocket expenditure made by those who had to visit health care centers. The expenditure on annual health care made by the
persons or households consists of the costs composed of three main categories, namely, medical cost, other direct cost and indirect cost. The medical costs contain mainly four items as doctor’s fees, costs of medicine, costs of diagnostic tests and other direct costs. Other direct costs include predominantly two items as transport costs and costs of diet. Indirect costs, which are also the important ingredients of the health care expenses, contain several items as interest costs on funds borrowed for treatment, loss of income of the ailing person, payment for the loss of earning of attending/caring person. The medical costs, other direct costs and indirect costs added together will give the total annual costs per household.

The out-of-pocket payments are determined by a number of factors, which may also be classified into the individual, households and community characteristics as before. In the spectrum of individual and household characteristics we incorporate the fact whether the persons or the households under consideration are protected by health insurance. The basic objective of such fact is to look into the impact of health insurance scheme on the expenditure made by the persons or households who took decision to visit health care centers. Other than the annual per capita income, sex, age, the region the persons of households belong to, literacy, frequency of illness, type of illness, severity of illness (number of days stayed in health care centers) and health consciousness are the important variables that we should include in the jurisdiction of the individual and household characteristics in order to consider the expenditure made in connection with the decision to visit health care centers. In the range of community characteristics religion and caste of the persons or households are important control variables.
3.2.2A. Analytical Model of Out-of-Pocket Payments for Health Care and Services

As an analytical model for the estimation of the expenditure of those who visited the health care centers, let us consider a log-linear function that would assess the out-of-pocket payments by the individuals.

*Out-of-Pocket Payments by the Households Provided that the Decision to Visit Health Care Centers is Positive* = \( f(\text{Income of the Household, Affiliation to Health Insurance Scheme, Householàs Characteristics, Community Characteristics and Random Disturbances}) \)

Therefore, we can specify the required model as

\[ \ln Y = X\beta + U \]

where

\( Y \) = out-of-pocket expenditure subject to the fact that the decision to visit health care centers by the persons or households is positive.

\( X \) = set of explanatory variables that are expected to affect the demand for health care and expenditure for the health care and services. It includes all the individual and household characteristics including the holding of health insurance scheme by the persons or households.
3.2.3. Decision to Purchase Health Insurance

In all the previous sections and sub-sections we have discussed the things that were responsible for any person to visit hospital and use hospital services and cares there. Subject to the fact that the individual concerned had visited hospital and used hospital services and cares there how his/her health care expenditure is affected by different factors, the important of which is definitely the affiliation of the person concerned to any health insurance scheme. It is expected that the health insurance affiliation would help the persons in reducing the burden of health care expenditure. Naturally the individuals should prefer to be protected by health insurance. Therefore, in this section we shall attempt to consider and analyze what factors are responsible for creating the demand for health insurance.

With this end in view, we have used a logit model that assesses how much more likely the people in one sort of category is to demand for health insurance compared to those who don’t belong to that category. We may also use probit model to this end. Since in our study in the district of Birbhum, we note that as a matter of health policy, health insurance scheme was not adopted by the government or by the employers as something compulsory, we cannot the policy effectiveness. Rather as general insurance scheme health insurance scheme is voluntary and people willing to be protected by health insurance are at liberty to purchase it from the open market. Since we would examine how the probability of purchasing health insurance of persons in one category differs
from that of the persons of other category, the logit or probit model is suitable.

In the determination of the demand for health insurance the individual characteristics of head of the family are important. Sex and different age groups of the head of the family will be included in this category of the characteristics affecting the demand for health insurance. In the household characteristics we include male literacy, female literacy, family size, worker population ratio, region where the person resides, health consciousness, illness ratio, awareness of the households, annual health care costs of the households, annual income/expenditure per household member, and premium benefit ratio as the important determinants of the demand for health insurance. The community characteristics of the households like religion and caste are also expected to affect the demand for health insurance.

3.2.3A. Analytical Model for Decision to Purchase Health Insurance

As the analytical model for the estimation of the decision to purchase health insurance, we shall use a binary logit model. A binary probit model can also be used. Whatever may be the type of model, logit or probit, it will be assumed that the decision to purchase health insurance is affected by different explanatory variables included in the category of the individual characteristics of the head of the family, the characteristics of the households and the characteristics of the community the households belong to.
Thus,

\[ \text{Decision to Purchase Health Insurance} = f(\text{Income of the Household, Characteristics of the Head of the Family, Households Characteristics, Community Characteristics and Random Disturbances}) \]

However, the analytical model for the decision to purchase health insurance can be developed in the same manner as we have developed the analytical model in the case of decision to visit health care centers already in section 3.2.1A.

3.2.4. Estimation of the Binary Response Models

The method of maximum likelihood is the most common way to estimate binary response models. Refer to equation (3.2.8), where in the binary response model, \( F(X', \beta) \) is the probability that \( Y_i = 1 \) and \( 1 - F(X', \beta) \) is the probability that \( Y_i = 0 \). Thus, if \( Y_i = 1 \), the contribution to the logarithm of the likelihood function for observation “i” is \( \ln F(X', \beta) \), while if \( Y_i = 0 \), the contribution is \( \ln \{1 - F(X', \beta)\} \). Since \( Y_i \) takes two values 0 and 1, the observed \( Y_i \) are just the realization of a binomial process with probability \( P_i = F(X', \beta) \). Given a sample of \( n \) observations on individual choices \( Y_i \), the first step towards maximum likelihood estimation of the unknown parameters \( \beta \) of both the Probit and the Logit models is to specify
the probability density functions of the observable random variables $Y_i$. They are

$$L(Y_i) = P_i^{y_i}(1 - P_i)^{1 - y_i} \quad (3.2.22)$$

Maximum likelihood estimation is based on the fact that the joint probability density function of the sample of $n$ independent observations is the product of the $n$ probability density functions $L(Y_i)$.

That is, we have

$$L(Y_1, Y_2, \ldots, Y_n) = \prod_{i=1}^{n} L(Y_i)$$

$$= \prod_{i=1}^{n} P_i^{y_i}(1 - P_i)^{1 - y_i}$$

$$= \prod_{i=1}^{n} [F(X'_i; \beta)^{y_i}][1 - F(X'_i; \beta)]^{1 - y_i} \quad (3.2.23)$$

If the parameters in $\beta$ vector were known, the joint probability density function could be used to calculate the probability that any set of $n$ choice outcomes occurs. However, $\beta$ is not known and after the sample is collected only the $n$ values of $Y_i$ and $X_i$ are known. The idea of maximum likelihood estimation is to choose (as estimates of $\beta$) the values of $\beta$ that maximize the probability of obtaining the sample that is actually observed. The resulting maximum likelihood estimates of the logit model are obtained by considering the joint probability density function given in equation
(3.2.23) to be a function of the unknown parameters \( \beta \), assuming that the sample outcomes \( Y_i \) and \( X_i \) are known.

The log likelihood function is given as follows.

\[
\ln L = \sum_{i=1}^{n} \left[ Y_i \ln F(X'_i, \beta) + (1 - Y_i) \ln \{1 - F(X'_i, \beta)\} \right] 
\]

\[
= \sum_{i=1}^{n} \left[ Y_i (\ln F(X'_i, \beta) - \ln \{1 - F(X'_i, \beta)\}) + \sum_{i=1}^{n} \ln \{1 - F(X'_i, \beta)\} \right] 
\]

\[
= \sum_{i=1}^{n} \left[ Y_i \ln \frac{F(X'_i, \beta)}{1 - F(X'_i, \beta)} + \ln \{1 - F(X'_i, \beta)\} \right] 
\]

\[
= \sum_{i=1}^{n} [Y_i X'_i - \ln \{1 + \exp(X'_i, \beta)\}] 
\]

\[
= \sum_{i=1}^{n} Y_i X'_i - \sum_{i=1}^{n} \{1 + \exp(X'_i, \beta)\} 
\]  

(3.2.24a)

(3.2.24)

In order to estimate \( \beta \) we differentiate (3.2.24) with respect to \( \beta \) to give normal equations as follows.

\[
\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} Y_i X'_i - \sum_{i=1}^{n} \{1 + \exp(X'_i, \beta)\}^{-1} \exp(X'_i, \beta). X_i = 0 
\]  

(3.2.25)

i.e., \[
\sum_{i=1}^{n} \left[ Y_i - \frac{\exp(X'_i, \beta)}{1 + \exp(X'_i, \beta)} \right] X_i = 0 
\]

i.e., \[
\sum_{i=1}^{n} (Y_i - P_i) X_i = 0 \quad [\text{from equation (3.2.17)}] 
\]
i.e., \[ \sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} P_i X_i = 0 \] (3.2.26)

Therefore, the Maximum Likelihood Estimation of \( \beta \) must satisfy the above condition.

Hence,

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} X_i \frac{\partial P_i}{\partial \beta'} - \sum_{i=1}^{n} X_i \frac{\exp(X_i' \beta)}{1 + \exp(X_i' \beta)} X_i' + \frac{[\exp(X_i' \beta)]^2}{[1 + \exp(X_i' \beta)]^2} X_i' \\
= -\sum_{i=1}^{n} X_i' P_i (1 - P_i) X_i' \tag{3.2.27}
\]

If "n" is greater than or equal to the number of elements in "X" and if the X's are not collinear the above matrix (3.2.27) will be negative definite. Using (3.2.27), we can apply Newton-Raphson method to maximize \( \ln L \).

Given the initial values of \( \beta \) in each iteration we use (3.2.17) to compute \( P_i \). Then, equations (3.2.16) and (3.2.17) can be used to compute \( \beta \) following the method of iteration by Newton-Raphson formula as

\[
\beta' = \beta^0 - k \left( \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right)^{-1} \left( \frac{\partial \ln L}{\partial \beta} \right) \tag{3.2.28}
\]
where, \( k \) is the step size of iteration and \( \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \) is the \( K \times K \) matrix of the second order partial derivatives of the log-likelihood function evaluated at the estimate \( \beta^0 \). The properties of the log-likelihood function (3.2.24) for both the normal and the logistic c.d.f.'s guarantee that this Newton-Raphson method of iteration will converge to the global maximum based on any set of starting values \( \hat{\beta}^0 \).

In order to start the iteration, we find out the initial estimate \( \beta^0 \) of \( \beta \) by regressing \( Y_i \) on the explanatory variable \( X_i \). The regression model, which is, \( Y_i = X_i' \beta + U_i \), is called the Linear Probability Model, since it estimates the probability of choosing a given alternative 1 by the linear function of \( X_i \). In this model the problem is that \( X_i' \beta \) which is \( P(Y_i = 1) \) may fall outside the (0,1) interval and the linear function of the continuous variable is not a good approximation to the probability of an event. Once the linear probability model gives the initial estimate \( \beta^0 \), the estimates of the logit or probit model can be found.

The maximum likelihood estimators are consistent, asymptotically efficient and asymptotically normally distributed. A consistent estimate of the asymptotic variance-covariance matrix of \( \hat{\beta} \) that can be used as a basis for hypothesis tests or confidence intervals is obtained by iterating with equation (3.2.28) and the corresponding result is
where \( p_i \) will be given by (3.2.6) with \( \beta \) replaced by \( \hat{\beta} \), the final set of estimates.

Given the procedure for obtaining estimates of the unknown parameters \( \beta \) of the Logit or Probit model, the question is whether the maximum likelihood estimation method has good statistical properties in repeated sampling situations. The properties of the maximum likelihood rule cannot, in general, be determined unless the sample size \( n \) is large. The resulting sampling properties are 'asymptotic' ones. When \( n \) is large, the maximum likelihood estimator \( \hat{\beta} \) for the models has a sampling distribution that is approximately normal with mean \( \beta \) and covariance matrix

\[
\text{cov}(\hat{\beta}) = (X'DX)^{-1}
\]

where \( X \) is the usual \((n \times k)\) design matrix of observations on \( k \) explanatory variables for \( n \) individuals, and \( D = \text{diag}(d_1, d_2, \ldots, d_n) \) is a diagonal matrix with elements

\[
d_i = \frac{[f(X'_i \beta)]^2}{F(X'_i \beta)[1 - F(X'_i \beta)]}
\]
where, \( f(X' \beta) \) and \( F(X' \beta) \) are the probability density function and cumulative distribution function of the standard normal random variable, respectively, evaluated at \( X' \beta \).

However, equation (3.2.24a) and equation (3.2.24) are the same likelihood function. The first order derivatives in maximizing the likelihood function are

\[
\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} Y_i \frac{f}{F} X_i - \sum_{i=1}^{n} (1 - Y_i) \frac{f}{1 - F} X_i
\]

(3.2.25a)

where \( F \) and \( f \) are respectively the relevant c.d.f.'s.

We may verify that for the Probit model, if we put

\[
f(t) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{t^2}{2} \right), \quad f'(t) = -tf(t) \quad \text{and} \quad F(-t) = 1 - F(t),
\]

then

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \left[ Y_i \frac{f + (X' \beta F)}{F^2} + (1 - Y_i) \frac{f - (X' \beta (1 - F))}{(1 - F)^2} \right] X_i X' \quad \text{(3.2.27a)}
\]
For logit model if we put

\[
F(t) = \frac{1}{1 + \exp(-t)}, \quad f(t) = \frac{\exp(-t)}{(1 + \exp(-t))^2}
\]

\[
1 - F(t) = \frac{\exp(-t)}{1 + \exp(-t)}, \quad \frac{f(t)}{F(t)} = 1 - F(t)
\]

and

\[
f'(t) = -f(t) \cdot F(t) \cdot (1 - \exp(t))
\]

then,

\[
\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} f(X_i'\beta)X_iX_i'
\]  \hspace{1cm} (3.2.27b)

Thus, in both the distribution \(\ln F\) is strictly concave. Hence, in these two standard cases, the maximum likelihood estimators may be obtained from the first order condition (3.2.25).

For the estimation of the log-linear model on the out-of-pocket payment for treatment, we shall use the ordinary least squares method.

3.2.5. **Tests of Hypotheses for the Binary Response Model**

In our binary response models we would like to test some hypotheses how the different characteristics, namely, individual, household and
community, affect the probability of decision to visit health care centers and decision to purchase health insurance. Since the characteristics are reflected through several explanatory variables, we would, in the context of our sample observations in the district of Birbhum, want to know the things as follows.

- How the explanatory variables that we have included in the probability models affect, individually or jointly, the probability of the decision to visit health care centers and the probability of the decision to purchase health insurance.

- In the context of the out-of-pocket expenditure, we would also test the hypotheses relating to the effect of the included explanatory variables on the expenditure made by the concerned persons or households on health care and services.

As we have already pointed out that the tests of the significance of individual parameters in our binary response models may be carried out as follows. In large samples, the maximum likelihood estimator \( \hat{\beta} \) from the probit and logit models has the distribution

\[
\hat{\beta} \sim N[\beta, \text{cov}(\hat{\beta})]
\]

where, \( \text{cov}(\hat{\beta}) \) is given by equation (3.2.30). Consequently,
Given the null and alternative hypotheses

\[
H_0 : \beta_m = 0 \quad \text{against} \quad H_1 : \beta_m \neq 0
\]  \hspace{1cm} (3.2.33)

Then, the t-ratio is

\[
t = \frac{\hat{\beta}_m}{\text{s.e.}(\hat{\beta}_m)}
\]  \hspace{1cm} (3.2.34)

If the null hypothesis is true, the t-ratio has a normal distribution (approximately) in large samples, and the critical values for the test may be taken from the standard normal distribution (if the sample is large) or the \(t_{(a-m)}\) distribution if the sample is not large.

Within the framework of maximum likelihood estimation, general and joint hypotheses about the parameter values may be tested in several ways.

Let the null and alternative hypotheses be

\[
\begin{align*}
H_0 &: R\beta = r \\
H_1 &: R\beta \neq r
\end{align*}
\]  \hspace{1cm} (3.2.35)
The linear equations represent \( J \) independent hypotheses about the parameters \( \beta \). When testing the significance of a single parameter, the basis for testing the null hypothesis in equation (3.2.33) is the sampling distribution of the maximum likelihood estimator for the Probit and Logit probability models,

\[
\hat{\beta} \sim N(\beta, \text{cov}(\hat{\beta}))
\]

Consequently,

\[
(R\hat{\beta} - r) \sim N((R\beta - r), R\text{cov}(\hat{\beta})R')
\]  \hspace{1cm} (3.2.36)

and if the hypotheses are true

\[
\lambda_w = (R\hat{\beta} - r)' [R\text{cov}(\hat{\beta})R']^{-1} (R\hat{\beta} - r) \sim \chi^2_J
\]  \hspace{1cm} (3.2.37)

where, \( J \) is the number of independent restrictions being tested. This chi-square test statistic is known as the Wald test statistic. It becomes large if the data do not support the null hypothesis, which is rejected if \( \lambda_w \geq \chi^2_J \).

The chi-square test statistic is appropriate for tests of joint hypothesis in the nonlinear probit and logit models rather than the usual \( F \)-statistic as there is no \( \sigma^2 \) term to deal with in the models. In small samples, however, the test statistic \( \lambda_w / J \) is sometimes advocated as a superior test statistic in
nonlinear models. If the null hypothesis is true, $\lambda_p / J$ is approximately distributed as $F_{(d, n - m)}$, and the $F$-test can be carried out in the usual way using this test statistic.

3.3. **Specification and Definition of the Variables**

Once we have considered the analytical framework of the models relating to several issues of demand for health insurance, we need to specify the actual models that are to be employed for the empirical estimation of the demand for health insurance. To this end, it is required, first of all, to specify and define the variables that had been included in our model of decision to visit health care centers/model of utilization of health care and services, model of out-of-pocket payments for health care and services and finally in the model of decision to purchase health insurance/model of participation in health insurance scheme. However, the variables in the study of demand for health insurance in the district of Birbhum are considered below according to the analytical models as discussed above. Some variables will be common in all the models. These common explanatory variables will be referred only if they are already defined and explained in any previous models.

*Model of Decision to Visit Health Care Centers*

This model actually explains the rate of utilization of health cares and services enjoyed by the people under consideration in the district of Birbhum during the year 2002–2003 the period for which we conducted this survey. We would focus on the determinants in the use of health care and services.
Decision to Visit Health Care Centers (DVHCC): It is a qualitative variable that is assigned the value 1 if the person takes the decision to visit health care centers; otherwise 0. Decision to visit health care centers is a qualitative dichotomous dependent variable. This variable actually indicates the rate of use of health care and services by the households. Thus, the positive decision to visit health care centers and utilization of health care and services are synonymous.

\[
DVHCC = \begin{cases} 
1, & \text{if the decision to visit health care centers is positive} \\
0, & \text{if the decision to visit health care centers is negative}
\end{cases}
\]

Explanatory Variables that Reflect the Individual and Household Characteristics

Decision to Purchase Health Insurance (DPHI): This is the most important explanatory variable that would affect the decision of the households or persons concerned to visit health care centers. Since the persons or households cannot predict beforehand the nature and type of disease and also the volume of costs required for the treatment, those who are protected by health insurance will not hesitate to take the decision to visit health care centers. But, it is not so easy due to cost of treatment for those who not protected by health insurance. However, the rate of utilization of health cares and services is likely to increase when the decision to purchase health insurance is positive. Since the decision to purchase health insurance is a
qualitative variable, we would assign the value ‘1’ if the individual/household is protected by health insurance and ‘0’ if not. Thus,

\[
D_{PHI} = \begin{cases} 
1, & \text{if the individual/household is protected by health insurance} \\
0, & \text{if the individual/household is not protected by health insurance}
\end{cases}
\]

**Sex:** Sex affects the rate of utilization of health cares and services and the decision to visit health care centers. It is a qualitative variable and takes two values, ‘1’ for male and ‘0’ for female.

\[
\text{Sex} = \begin{cases} 
1, & \text{if the person is male} \\
0, & \text{if the person is female}
\end{cases}
\]

**Age Group:** Age is, no doubt, likely to affect the decision to visit health care centers and the rate of utilization of health cares and services. Age has been measured in years as usual. It is natural to think that the persons of old age are more vulnerable to disease than the persons of young and middle age. Therefore, we have considered it rational that not a single age group should be incorporated in the analysis of determining the decision to visit health care centers or the rate of utilization of health cares and services. We have divided the persons in three age groups; namely,
• age group 1 where all persons with age below 25 years have been considered;
• age group 2 where all persons in the middle age in between 25 years and 50 years have been considered; and
• age group 3 where all persons of old age with age above 50 years have been considered.

We assign '1' if the person in question belongs to any particular age group and '0' if not, as follows.

\[
\text{Age Group 1} = \begin{cases} 
1, & \text{if the person belongs to the young age group below 25 years} \\
0, & \text{if the person does not belong to the young age group below 25 years}
\end{cases}
\]

\[
\text{Age Group 2} = \begin{cases} 
1, & \text{if the person belongs to the middle age group between 25 and 50 years} \\
0, & \text{if the person does not belong to the middle age group between 25 and 50 years}
\end{cases}
\]
Age Group 3 = \[
\begin{cases}
1, \text{ if the person belongs to the old age group with age above 50 years} \\
0, \text{ if the person does not belong to the old age group with age above 50 years}
\end{cases}
\]

**Region:** The residential background (rural or urban) of the person is important in the determination of the rate of utilization of health cares and services and the decision to visit health care centers. Most of the health care centers are within the easy access of the urban people. But this is not true for the rural people. Further, in most of the rural areas, the quality of medical cares and services rendered by the health care centers is very much poor in quality. Naturally, the rural people, unless there is some emergency are very much reluctant to visit the health care centers. The road, transportation in the rural area and other system of communication in between rural areas and health care centers in the rural areas are not also in the state of good conditions, as we have noted in the district of Birbhum, that can persuade the rural individuals to take prompt decision to visit health care centers. However, region as one of the determinants of the rate of utilization of health cares and services is a qualitative variable that takes the value ‘1’ if the person who takes the decision to visit health care centers and uses health cares and services, resides in the urban area, and ‘0’ if the person resides in the rural area.
Region = \begin{cases} 
1, \text{ if the person or household resides in the urban area} \\
0, \text{ if the person or household resides in the rural area}
\end{cases}

Literacy: Literacy is important in the decision of visiting health care centers and using health cares and services. We have considered a person literate and educated if he/she has been taught at least ten years in school and passed the school leaving examination which is called school final examination or Madhyamik Pariksha in West Bengal. Therefore, instead of measuring the literacy by the number of years of attending the educational institutes, we have considered the literacy as dummy independent variable and it is assigned the value ‘1’ if the person has passed (which is considered to be the minimum level of education) School Final examination or some higher examination and ‘0’, if not.

\begin{align*}
\text{Literacy} = \begin{cases} 
1, \text{ if the person has passed the school final examination or higher examination.} \\
0, \text{ if the person has not passed even the school final examination}
\end{cases}
\end{align*}

Frequency of Illness: Frequency of illness has been defined as the number of times in a year the person became ill. It is measured simply in numerals. The
more the number of times one becomes ill, the more likely is the rate of utilization of health cares and services and the decision to visit health care centers. In the district of Birbhum, we have noted and perhaps it is true everywhere, that even if the frequency of illness is high for those who are not protected by health insurance, the rate of utilization of health cares and services is low.

Health Consciousness: We have introduced this variable in our empirical study since the health consciousness is important for a healthy life and it surely increases the rate of utilization of health cares and services and affects the decision to visit health care centers. The measurement of health consciousness apparently comes to us problematic. But, we have tried to solve in our own way. We have set a number of commonsensical criteria of which not all but a few are fulfilled by the person we interviewed in our sample in the district of Birbhum. For example, suppose we have set ten commonsensical criteria and the criterion like the immunization of the babies is compulsory. Now the degree of health consciousness is measured by the ratio of the number of commonsensical criteria fulfilled by the persons or households to the total number of commonsensical criteria, as set by us, provided that the task of immunization of the babies is not ignored/postponed/cancelled/neglected due to any sort of superstitions/dictum/misconceptions/taboos. If the households fulfill all the commonsensical criteria but the immunization of the babies, the degree of health consciousness in such case has been measured to be nil or zero. But, if five or six commonsensical criteria along with the immunization of

---

1 See section 3.3.1. for the commonsensical criteria

154
the babies are satisfied, the degree of health consciousness in such case has been measured as 50% or 60% respectively. We could have retained such type of measurement of health consciousness. But, instead of that we have considered the health consciousness variable as qualitative variable that takes the value ‘1’ if the degree of health consciousness is 50% or above and ‘0’ if not. Thus, a person of household is said to be health conscious if the degree of health consciousness is above 50% or above.

Health Conscious =
\[
\begin{cases} 
1, & \text{if the degree of health consciousness is above 50% or above} \\
0, & \text{if the degree of health consciousness is below 50%}
\end{cases}
\]

Annual Income per Capita: It is most important of all explanatory variables affecting the decision of the households to visit health care centers and rate of utilization of health cares and services. Annual income is measured as the average expenditure of a household per annum and per member of each household. The unit of measurement of income is rupee. While collecting data for this study in the district of Birbhum, we had in our mind a doubt and also a fear that the respondents may not correctly reply in response to our enquiry of their income. As the respondents are engaged in different types of jobs relating to agriculture, casual labour, self employment and services in the organized sector, there was a always possibility of getting an over-estimated or under-estimated figures on their income. In the situation, mainly related to the casual labour, where a part of wage is paid in kind and
the other part in cash, income calculation is rather difficult in the sense that it needs the knowledge about the-then price level also. In order to escape such a problem, we have thought it to be a rational approach to consider as the determinant of the decision to visit health care centers and rate of utilization of health cares and services the actual expenditure spent by the household for its each member.

Further the picture of the decision to visit health care centers and the rate of utilization of health cares and services by the households or persons is likely to be overshadowed and bound to provide wrong signal when we consider the level of income of all households as the explanatory variable, without emphasizing on the fact of income differentials of these households. In such a case, some sort of average picture will be revealed. The results of the households who belong to the lower income category will be erroneous and mis-reported due to the presence of the households of other income categories. So is true for the results of other income categories due to the presence of households of lower income category. Therefore, what we have suggested is to divide the households in three income categories, namely, lower, middle and higher income groups.

- The households with annual income below Rs.36000 belong to the lower income category.
- The households with annual income between Rs.36000 and Rs.80000 belong to the middle income category.
- The households with annual income Rs.80000 and above belong to the higher income category.
We would like to look into the effect on the decision to visit health care centers and rate of utilization of health cares and services for these three income group households. We assign the value ‘1’ if the household belongs to any particular income category and ‘0’ otherwise. Thus, we have

\[
\begin{align*}
\text{Lower Group} &= \begin{cases} 
1, & \text{if the household belongs to the group with income} \\
& \text{below Rs.36000 per annum} \\
0, & \text{otherwise}
\end{cases} \\
\text{Middle Group} &= \begin{cases} 
1, & \text{if the household belongs to the group with income} \\
& \text{between Rs.36000 and Rs.80000 per annum} \\
0, & \text{otherwise}
\end{cases} \\
\text{Higher Group} &= \begin{cases} 
1, & \text{if the household belongs to the group with income} \\
& \text{Rs.80000 and above per annum} \\
0, & \text{otherwise}
\end{cases}
\end{align*}
\]
Explanatory Variables that Reflect the Community Characteristics

Religion: Religion is a qualitative variable affecting the decision to visit health care centers and the rate of utilization of health cares and services. We have considered households or persons who are divided into the community of the Hindu and non-Hindu. If the respondent belongs to the Hindu community we assign the value ‘1’ and ‘0’ otherwise. Thus,

\[
\text{Religion} = \begin{cases} 
1, & \text{if the household or the person belongs to the Hindu community} \\
0, & \text{otherwise}
\end{cases}
\]

Caste: Caste has got an important role in the determination of the decision to visit health care centers and rate of utilization of health cares and services. We have considered the households of different castes as divided into four broad categories, namely, General Caste, Scheduled Caste, Scheduled Tribe and Other Backward Community. Caste is a dummy variable and takes the value ‘1’ if the person or the household belongs to any particular category and ‘0’ otherwise.
\[
\text{General Caste} = \begin{cases} 
1, & \text{if the household or the person belongs to the General} \\
0, & \text{otherwise}
\end{cases}
\]

\[
\text{Scheduled Caste} = \begin{cases} 
1, & \text{if the household or the person belongs to the} \\
0, & \text{otherwise}
\end{cases}
\]

\[
\text{Scheduled Tribe} = \begin{cases} 
1, & \text{if the household or the person belongs to the} \\
0, & \text{otherwise}
\end{cases}
\]

\[
\text{OBC} = \begin{cases} 
1, & \text{if the household or the person belongs to the OBC} \\
0, & \text{otherwise}
\end{cases}
\]
Subject to the condition that the person or the household takes the decision to visit health care centers and gets involved in the health care use, we would like to examine the impact of several determinants of out-of-pocket expenditure for health care use during the Year 2002–2003 in the district of Birbhum. Thus the out-of-pocket payment of those who visited the health care centers and used health cares and services is the dependent variable in this model. The out-of-pocket payment is explained by a number of variables as follows.

Explanatory Variables that Reflect the Individual and Household Characteristics

The out-of-pocket payment is determined by the independent variables as specified below. The list of the explanatory variables includes the decision to purchase health insurance (1 = yes), Sex (1 = male), Age Group 1 (Age < 25), Age Group 2 (Age = 25–50), Age Group 3 (Age > 50), Region (1 = urban), Literacy (1 = S.F. or above), Frequency of Illness, Type of Illness (1 = Complicated), Severity of Illness (number of days stayed in health care centers), Health Conscious (1 = yes), Annual Income per Capita, Lower Group (1 = yes), Middle Group (1 = yes) and Upper Group (1 = yes).
Explanatory Variables that Reflect the Community Characteristics

As before the list of explanatory variables that belongs to the community characteristics and affects the out-of-pocket payments includes Religion (1 = Hindu), Caste (1 = General Caste), Caste (1 = Scheduled Caste), Caste (1 = Scheduled Tribe) and Caste (1 = OBC).

We have explained all the explanatory variables above except the Type of Illness (1 = Complicated), Severity of Illness (number of days stayed in health care centers), which are two new variables in the category of individual and household characteristics.

**Type of Illness**: The out-of-pocket expenditure made by the household will vary depending on the nature and type of illness. If the type of illness is complicated, it requires much cost. Thus the type of illness is a dummy variable, which may also be considered as a control variable in the sense that it distinguishes between low-cost versus high-cost diseases. We assign the value ‘1’ if the type illness is complicated and ‘0’ if not. Thus,

\[
\text{Type of Illness} = \begin{cases} 
1, & \text{if the type of illness is complicated} \\
0, & \text{otherwise}
\end{cases}
\]

**Severity of Illness**: Severity of illness will definitely affect the out-of-pocket payment for using health care and services. We measure the severity of illness by the number of days the person will have to stay in health care
centers. It is natural that if the illness is complicated and severe, it needs a long stay in the health care centers and leading to an increase in the volume of the out-of-pocket expenditure. By long stay consider a stay in the health care centers for more than three days. Normally, the patient after his or her admission in the health care center may be released on or before the third day if the disease is not severe. We assign the value ‘1’ if the illness is severe, otherwise ‘0’. Hence,

\[
\text{Severity of Illness} = \begin{cases} 
1, & \text{if the illness is severe and it needs stay in the health care centers for more than three days} \\
0, & \text{otherwise}
\end{cases}
\]

**Model of Decision to Purchase Health Insurance**

The model of decision to purchase health insurance actually explains the determinants of demand for health insurance. In such case, the decision to purchase health insurance is the qualitative dependent variable that takes two values ‘1’ and ‘0’ if the individual purchases and does not purchase the health insurance. The decision to purchase health insurance as the dependent variable has been explained and defined in the beginning of this section.
Explanatory Variables that Reflect the Individual Characteristics of Head of the Family

The decision to purchase health insurance is determined by the independent variables as specified below. The list of the explanatory variables that reflects individual characteristics of head of the family includes Sex (1 = male), Age Group 1 (Age < 25), Age Group 2 (Age = 25–50) and Age Group 3 (Age > 50).

Explanatory Variables that Reflect the Household Characteristics

Other than the explanatory variables reflecting the individual characteristics of the head of the family, we should also consider the explanatory variables that would reflect the characteristics of the households, in the determination of the decision to purchase health insurance. Of the explanatory variables exhibiting the household characteristics, the most important are Male Literacy (1 = S.F. or above), Female Literacy (1 = S.F. or above), Family Size, Worker Population Ratio, Health Consciousness (1 = yes), Region (1 = urban), Illness Ratio, Awareness (1 = yes), Annual Health Care Costs, Annual Income per Household Member and Premium Benefit Ratio.

Male Literacy and Female Literacy: In the determination of rate of utilization of health cares and services and out-of-pocket expenditure made
by the households we have defined and specified literacy as one of the explanatory variables. But, we have considered male literacy and female literacy as two separate explanatory variables in the determination of decision to purchase health insurance. The justification of such classification of literacy lies in the fact that if female becomes literate and educated, they become empowered to exert their opinion and influence much more the decision of the family in the purchase of health insurance than in a situation where females are not literate. However, both male and female literacy are considered to be dummy variable as follows.

\[
\text{Male Literacy} = \begin{cases} 
1, & \text{if the male has passed the school final examination or higher examination.} \\
0, & \text{if the male has not passed even the school final examination}
\end{cases}
\]

\[
\text{Female Literacy} = \begin{cases} 
1, & \text{if the female has passed the school final examination or higher examination.} \\
0, & \text{if the female has not passed even the school final examination}
\end{cases}
\]

164
Family Size: Family size likely to affect the decision to purchase health insurance. Family size is measured in number of family members.

Worker Population Ratio: The number of worker as proportion to family size is defined as the worker population ratio. Then the worker population ratio subtracted from one gives us the dependency ratio in a family. The worker population ratio or dependency ratio measured in pure number, will affect the decision to purchase health insurance.

Illness Ratio: We have considered the frequency of illness for a person in any household as the explanatory variable in the model of decision to visit health care centers and rate of utilization of health cares and services and also in the model of out-of-pocket expenditure for health care use. We would modify that variable slightly in the present situation. We specify the modified variable as "Illness Ratio". In order to calculate "Illness Ratio" we simply consider the number of cases of illness per family (not for a person but for all members of the household) divided by the family size during the last one year prior to the period (2002 – 2003) of collection of this sample.

Awareness: If the individual is well aware of the terms and conditions of the health insurance schemes, it helps taking the decision to purchase health insurance. Therefore, awareness, which is a qualitative variable is important. We assign the value ‘1’ if the person of household is well aware of the health insurance scheme, and ‘0’ if not. Thus,
Awareness = \begin{cases} 
1, & \text{if the person or household is well aware of the health insurance scheme} \\
0, & \text{if the person or household is not well aware of the health insurance scheme} 
\end{cases}

**Annual Health Care Cost per Household Member:** The expenditure on annual health care made by the persons or households consists of the costs composed of three main categories, namely, medical cost, other direct cost and indirect cost. The medical costs contain mainly four items as doctor's fees, cost of medicines, cost of diagnostic tests and other direct costs. Other direct costs include predominantly two items as transport costs and costs of diet. Indirect costs, which are also the important ingredients of the health care expenses, contain several items as interest costs on funds borrowed for treatment, loss of income of the ailing person, payment for the loss of earning of attending/caring person. The medical costs, other direct costs and indirect costs added together will give the total annual costs per household. The annual health care cost has been considered per household member.

**Annual Income per Household Member:** We have defined and specified this variable already in previous two models.

**Premium Benefit Ratio:** Premium refers to the prepayment for health insurance scheme to ensure some future benefits in terms of coverage of cost of treatment. The ratio of premium to benefit is a very much important
variable in the determination of decision to purchase health insurance. The higher the prepayment required ensuring future benefit for treatment, the lower will be the willingness to go in for purchasing health insurance scheme.

**Explanatory Variables that Reflect the Community Characteristics**

The decision to purchase health insurance is also governed by the list of factors like Religion (1 = Hindu), Caste (1 = General Caste), Caste (1 = Scheduled Caste), Caste (1 = Scheduled Tribe) and Caste (1 = OBC).

3.3.1. Commonsensical Criteria of Health Consciousness

In order to measure the degree of health consciousness of the persons we interviewed in the district of Birbhum, we asked the head of the family a set of questions relating to their life style in maintaining good health. The set of questions may not be of full proof of health consciousness. There may be other questions, which may capture better the idea that will reveal the health consciousness. But, as a first hand measure, the questions we use for measuring the health consciousness will also reveal some degree of health consciousness. The set of questions asked to head of the family is as follows.

**Questions for babies below five years**

(i) Have you followed for your babies since their birth the process of immunization?
Has any vaccination in the process of immunization or intake of pulse polio as arranged by the health department, for your babies below five years been dropped, forgotten or neglected?

Has any vaccination in the process of immunization for your babies deferred? If deferred, for which reason?

Do you maintain all sorts of hygienic matters for your babies like cleansing of clothes, washing, boiling of water for its drinking etc.?

Do you maintain food requirement as per age, weight and height of your babies?

Do you follow any programme of routine check up for your babies and check its weight and height?

Do you consult any child specialist when your baby is ill?

Do you consult, when your baby is ill, any doctor who is not at all a child specialist?

Questions for family members other than babies below five years

Do you and your family members use sanitary latrine?

Do you participate in health education programme?

Do you pursue family planning programme?

In case of suffering from illness, do you go to gunin, tantrika, ojha etc? Or simply have you any faith in the deities that would come for your help in rescuing from the disease?

Are you smoker?

Are you addicted to alcohol?
(vii) Do you maintain food as per calorie requirement of 3000 to 3500 per day?

(viii) When you are ill, do you go to a doctor?

(ix) Do you, when you go to a doctor, follow doctor’s advice?

(x) Do you take the medicine in its full course as prescribed by the doctor?

(xi) Or, do you stop taking medicine before the completion of the course, if you feel better after the intake of initial doses of medicines?

(xii) Do you visit the doctor for follow-up advice or drop going there since you are okay after the first visit?

(xiii) As now a days it is seen everywhere, do you go in for self-medication?

(xiv) Is it that you go to a medicinal shop and purchase the medicine on telling your problem to staff who run the medicinal shop?

(xv) Do you neglect any physical suffering simply on self-consolation?

(xvi) Are you hypochondriac?

These are the questions asked to the head of the family for deciding the health consciousness matter of the persons/households we interviewed in the district of Birbhum.

3.4. Specification of the Econometric Model for Health insurance

We intend in this section to develop the model for different issues relating to the demand for health insurance. Let us consider three such issues
in connection with the demand for health insurance. First, we would consider the impact of health insurance on the probability of visiting health care centers. By health care centers we mean hospitals, which fall in the government and private sector. The treatment in the hospitals or the primary health centers run by government was almost free of cost a few years back. This is not so today. However, even if some cost is required, as now a days some costs are also stipulated in the public sector hospitals, it is not so exorbitantly as high as that in the private sector hospitals. But, since the availability and the standard of medical services supplied in the government hospitals is poor in quality, the alternative for the insurance holder is to enjoy the facility of medical services, even if its price is high, supplied in the health care centers run by the private sector. Second, in going to the health care center the people would need to spend for purchasing health cares and services and they will have to shoulder the burden of expenditure out of their pocket. We would like to see how this out of pocket expenditure is protected (affected) when the individual is (is not) the holder of health insurance scheme. Third and final, if the probability of visiting health care center is positive, the individual will have to shoulder the cost of using the health care services. The costs of health care services may be financed out of own pocket in the absence of any health insurance in the ownership of the incumbent or the health care expenses may be met through the facility of health insurance scheme. Therefore, it is justified that the demand for the health insurance be considered.
3.4.1. A Logit Model for Visiting Health Care Centers

The development of our logit model for visiting the health care centers that would assess the rate of utilization of health cares and services in the district of Birbhum will follow the line as we have developed the general analytical framework in section 3.2 of this chapter. Therefore, we specify the model as follows.

**Model A: Logit Model When Income is a Continuous Variable**

\[
\text{Decision to visit health care centers or Rate of utilization of health cares and services} = \text{Constant} \\
+ \alpha_1 \text{decision to purchase health insurance} + \alpha_2 \text{sex} + \\
\alpha_3 \text{age group}1 + \alpha_4 \text{age group}2 + \alpha_5 \text{age group}3 + \alpha_6 \text{region} + \alpha_7 \text{literacy} + \alpha_8 \text{frequency of illness} + \\
\alpha_9 \text{health consciousness} + \alpha_{10} \ln(\text{annual income per capita}) + \alpha_{11} \text{religion} + \alpha_{12} \text{general caste} + \\
\alpha_{13} \text{scheduled caste} + \alpha_{14} \text{scheduled tribe} + \alpha_{15} \text{other backward community} + \text{error}
\]

(3.4.1A)
**Model B: Logit Model When Dummies are Used for Income Groups**

\[
\text{Decision to visit health care centers or Rate of utilization of health cares and services} = \text{Constant} \\
+ \alpha_1 \text{decision to purchase health insurance} + \alpha_2 \text{sex} + \alpha_3 \text{age group1} + \alpha_4 \text{age group2} + \alpha_5 \text{age group3} + \alpha_6 \text{region} + \alpha_7 \text{literacy} + \alpha_8 \text{frequency of illness} + \alpha_9 \text{health consciousness} + \alpha_{10} \text{lower income group} + \alpha_{11} \text{middle income group} + \alpha_{12} \text{upper income group} + \alpha_{13} \text{religion} + \alpha_{14} \text{general caste} + \alpha_{15} \text{scheduled caste} + \alpha_{16} \text{scheduled tribe} + \alpha_{17} \text{other backward community} + \text{error}
\]

(3.4.1B)

where decision to visit health care centers or Rate of utilization of health cares and services is the qualitative dependent variable taking two values ‘1’ in the case positive response and ‘0’ in the case of negative response.

**3.4.2. A Log Linear Model for Out of Pocket Expenditure**

The model for out-of-pocket expenditure incurred by the households who visit the health care centers is semi-logarithmic and is specified as follows. We have considered two models, model A and model B. Model A considers annual income per household member as the explanatory variable along with other explanatory variables and model B considers along with other explanatory variables, the income group dummy variables.

172
Model A: Log-linear Model When Income is a Continuous Variable

\[
\ln(\text{out-of-pocket payment} \mid \text{visit} > 0) = \text{Constant} + \\
\beta_1 \text{decision to purchase health insurance} + \beta_2 \text{sex} + \beta_3 \text{age group1} + \beta_4 \text{age group2} + \beta_5 \text{age group3} + \beta_6 \text{region} + \\
\beta_7 \text{literacy} + \beta_8 \text{frequency of illness} + \beta_9 \text{type of illness} + \\
\beta_{10} \text{severity of illness} + \beta_{11} \text{health consciousness} + \\
\beta_{12} \ln(\text{annual income per capita}) + \beta_{13} \text{religion} + \beta_{14} \text{general caste} + \beta_{15} \text{scheduled caste} + \beta_{16} \text{scheduled tribe} + \beta_{17} \text{other backward community} + \text{error} \quad (3.4.2A)
\]

Model B: Log-linear Model When Dummies are Used for Income Groups

\[
\ln(\text{out-of-pocket payment} \mid \text{visit} > 0) = \text{Constant} + \\
+ \beta_1 \text{decision to purchase health insurance} + \beta_2 \text{sex} + \\
\beta_3 \text{age group1} + \beta_4 \text{age group2} + \beta_5 \text{age group3} + \\
\beta_6 \text{region} + \beta_7 \text{literacy} + \beta_8 \text{frequency of illness} + \\
\beta_9 \text{type of illness} + \beta_{10} \text{severity of illness} + \beta_{11} \text{health consciousness} + \\
\beta_{12} \text{lower income group} + \\
\beta_{13} \text{middle income group} + \beta_{14} \text{upper group} + \\
\beta_{15} \text{religion} + \beta_{16} \text{general caste} + \beta_{17} \text{scheduled caste} + \\
\beta_{18} \text{scheduled tribe} + \beta_{19} \text{other backward community} + \\
\text{error} \quad (3.4.2B)
\]
3.4.3. A Probabilistic (Logit or Probit) Model for Demand for Health Insurance

In the present sub-section we are going to explain factors that are responsible for affecting the demand for health insurance. To this end we may specify the logit or the probit model by which the probability of demand for health insurance would be assessed. Whatever model may be chosen, the basic issue to consider is how the decision to purchase health insurance is affected by the relevant explanatory variables as specified in a previous sub-section. We are going to present the probabilistic model for demand for health insurance.

\[
\text{Decision to purchase health insurance} = \text{Constant} + \gamma_1 \text{sex} + \gamma_2 \text{age group1} + \gamma_3 \text{age group2} + \gamma_4 \text{age group3} + \gamma_5 \text{male literacy} + \gamma_6 \text{female literacy} + \gamma_7 \text{family size} + \gamma_8 \text{worker population ratio} + \gamma_9 \text{region} + \gamma_{10} \text{health consciousness} + \gamma_{11} \text{illness ratio} + \gamma_{12} \text{awareness} + \gamma_{13} \text{annual health care cost per capita} + \gamma_{14} \ln(\text{annual income per capita}) + \gamma_{15} \text{premium benefit ratio} + \gamma_{16} \text{religion} + \gamma_{17} \text{general caste} + \gamma_{18} \text{scheduled caste} + \gamma_{19} \text{scheduled tribe} + \gamma_{20} \text{other backward community} + \text{error} \quad (3.4.3)
\]

where decision to purchase health insurance is the qualitative dependent variable taking two values ‘1’ in the case positive response and ‘0’ in the case of negative response.
The equation (3.4.3) would be considered for assessing the probability of decision to purchase health insurance in five models.

- **Model 1:** It's a pooled model that takes all the households (for whom we collected data for this study during 2002 – 2003) in the district of Birbhum under consideration irrespective of the type of occupation they are involved in, for assessing the probability of decision to purchase health insurance. This model, therefore, is not capable of looking at and looking into the differences in respect of creating demand for health insurance by people of alternative category of occupation. Rather some sort of average picture of demand for health insurance is captured in such pooled model.

- **Model 2:** It considers the demand for health insurance for the service holders. In other words, this model will provide us with a picture of demand for health insurance of the households in the organized sectors.

- **Model 3:** It is designed to assess the probability of decision to purchase health insurance of the cultivators. This model will show how the agricultural sector in the district of Birbhum accepts and responds to the issue of health insurance.
• Model 4: Apart from the agricultural and organized sector in creating the demand for health insurance, we would like to examine how the casual labour who has got no fixed and regular source of income take the decision to purchase health insurance. This section of population is very poor in relation to the other sections of the society.

• Model 5: It's a model for demand for health insurance of the businessmen, who are no doubt the financially sound people of the society.

3.4.4. Endogeneity and Selection Bias

It is clear from the above discussion that the health insurances have definite impact on the health care use and expenditure. Naturally, the researchers in such a situation would face two types of problems. The first one is the problem of "endogeneity" and the second the problem of "self-selection". According to Waters (1999), in the study of the impact of health insurance on the health care use, the potential endogeneity of the choice of insurance for health care use is the main problem and it creates some sort of potential selection bias. In this case, there is a possibility of adverse selection for the individuals who of their own choose the insurance program. The choice of the insurance schemes may have some unobservable features "that might make it more likely for them to join the program, and might also influence their decision to use health care services" (see Jutttings, 2003). An observed relation between the health insurance affiliation and health care use
and expenditure due to unobservable characteristics may therefore lead to the overestimation of the insurance effect.

In the sphere of development economics the issue of challenge between "endogeneity" and "self-selection" is very much common. For example, there are different studies (Coleman, 1999; Nada, 1999) that deal with the measurement of the effect of micro finance institutions. Some studies (Bedi and Gaston, 1999) have focused on measuring the return of education through the impact of education on economy. Some studies by Waters (1999) and Yip and Berman, (2001) have concentrated on the analysis of the effect of health insurance on health care demand and financial protection. These studies have shown the problem of the difficulty of randomly assigning some individuals to nonprogram control and some others to program treatment groups. "It flows from this that the identification of an adequate control group is the first and most important step when trying to control for self-selection" (Juttings, 2003).

In order to test the endogeneity/exogeneity of the variable "the decision to purchase health insurance" we can follow a procedure as used by Waters (1999). Waters' approach is designed to test the significance of residuals or predicted values from the secondary equation when inserted in the primary equation (see Smith and Blundell, 1986). Therefore, the logit/probit equation of decision to purchase health insurance that we shall use for estimation is the secondary equation and the decision to visit health care centers/the equation of the rate of utilization of health cares and services is the primary equation. The other primary equation is the equation

177
of the logarithm of out-of-pocket payment for health cares and services. However, the approach by Waters (1999) is described below.

- We estimate a reduced form of the variable “the decision to purchase health insurance”, the endogeneity/exogeneity of which is our chief concern and to be tested.

- From the estimated reduced form equation the predicted values of the decision to purchase health insurance are to be computed.

- The predicted and the actual observed values of the decision to purchase health insurance are to be then inserted in the equation of decision to visit health care centers/rate of utilization of health cares and services and the equation of out-of-pocket payment for health cares and services.

- We then test the hypothesis of non-zero coefficient of the predicted value of the decision to purchase health insurance in the primary equations.

- If the hypothesis of non-zero coefficient of the predicted value of the decision to purchase health insurance in the primary equations appears to be statistically significant, we infer that the decision to purchase health insurance or the participation in the insurance programme is an endogenous variable.
In actuality the test procedure described above is similar to the Hausman test of testing the significance of an omitted variable. Intuitively, this test examines whether the unobserved variables in the reduced form equation help to explain the variation in decision to visit health care centers or rate of utilization of health cares and services after controlling for the observed explanatory variables (see, Bollen et. el. 1995). If the coefficient of the predicted value term included in the primary equations, is significantly different from zero, we conclude that the unobserved variables affect the predicted value and it must have some correlation with the error term in the primary equations. Since, our secondary equation is binary probabilistic model, we can use the residual of this secondary equations and this is tantamount to using the predicted values. The primary equations, which will be used to test the exogeneity/endogeneity of the variable ‘decision to purchase health insurance’, are given below.

\[
\text{Decision to visit health care centers or Rate of utilization of health cares and services} = \text{Constant} \\
+ \alpha_{1\text{actual}} \text{ actual decision to purchase health insurance} + \\
\alpha_{1\text{predicted}} \text{ predicted decision to purchase health insurance} + \alpha_2 \text{ sex} + \alpha_3 \text{ age group1} + \alpha_4 \text{ age group2} + \\
\alpha_5 \text{ age group3} + \alpha_6 \text{ region} + \alpha_7 \text{ literacy} + \alpha_8 \text{ frequency of illness} + \alpha_9 \text{ health consciousness} + \alpha_{10} \ln(\text{annual income per capita}) + \alpha_{11} \text{ religion} + \alpha_{12} \text{ general caste} + \\
\alpha_{13} \text{ scheduled caste} + \alpha_{14} \text{ scheduled tribe} + \alpha_{15} \text{ other backward community} + \text{error}
\]  

(3.4.4)
\[ \ln(\text{out-of-pocket payment} | \text{visit} > 0) = \text{Constant} + \]
\[ \beta_1 \text{actual decision to purchase health insurance} + \]
\[ \beta_1 \text{predicted decision to purchase health insurance} + \beta_2 \text{sex} + \beta_3 \text{age group1} + \beta_4 \text{age group2} + \beta_5 \text{age group3} + \beta_6 \text{region} + \beta_7 \text{literacy} + \beta_8 \text{frequency of illness} + \beta_9 \text{type of illness} + \beta_{10} \text{severity of illness} + \beta_{11} \text{health consciousness} + \beta_{12} \ln(\text{annual income per capita}) + \beta_{13} \text{religion} + \beta_{14} \text{general caste} + \beta_{15} \text{scheduled caste} + \beta_{16} \text{scheduled tribe} + \beta_{17} \text{other backward community} + \text{error} \]

(3.4.5)

The primary equation (3.4.4) is actually the equation (3.4.1A) with the predicted decision to purchase health insurance as the additional explanatory included in the primary equation of the decision to visit health care centers/rate of utilization of health cares and services. Similarly the primary equation (3.4.5) is actually the equation (3.4.2A) with the predicted decision to purchase health insurance as the additional explanatory included in the primary equation of the log of out-of-pocket payment for visiting the health care centers and utilization of health cares and services there.

What we would like to do is simply to carry out the test of significance of the predicted decision to purchase health insurance as an
explanatory variable in equations (3.4.4) and (3.4.5). That is, we would test the null hypothesis that $H_0: \alpha_{1,\text{predicted}} = 0$ against the alternative hypothesis that $H_1: \alpha_{1,\text{predicted}} \neq 0$ in equation (3.4.4). Similarly, we would test the null hypothesis that $H_0: \beta_{1,\text{predicted}} = 0$ against the alternative hypothesis that $H_0: \beta_{1,\text{predicted}} \neq 0$. If the null hypothesis is accepted, we conclude that the variable 'decision to purchase health insurance' is an exogenous variable; otherwise it is an endogenous variable.

3.5. Specification of Hypothesis

Once we have defined and specified the model of demand for health insurance that will assess the probability of the decision to purchase health insurance, it is our task to set the hypotheses that are to be tested in the light of the sample collected in the district of Birbhum during the period 2002 – 2003.

3.5.1. Hypothesis Relating to the Model for Visiting Health Care Centers

A number of hypotheses that would explain how the decision to visit health care centers or the rate of utilization of health cares and services is affected by the different variables may be stated as follows. **The logit model that we have used will enable us to test how the several explanatory variables affect the probability that a person will visit hospital (health care centers) and use hospital cares and services.** Therefore let us consider Model A and Model B in section 3.4.1.
Hypothesis 1

Persons protected by health insurance are much more likely to visit hospital (health care centers) and use health cares and services than those not protected by health insurance. The probability of using hospital cares and services for persons protected by health insurance increases contrasted with the probability of the persons not protected by health insurance. That is, $\alpha_1 > 0$ (refer to equation (3.4.1)).

Hypothesis 2

The males are much less likely to visit hospital and use hospital cares and services in contrast to females. The probability of using hospital care and services for males in contrast to that of females is low. That is, $\alpha_2 < 0$ (refer to equation (3.4.1)).

Hypothesis 3

Persons in the age group below 25 years are much less likely to visit hospital and use hospital cares and services in contrast to persons of other age groups. The probability of using hospital cares and services for persons in the age group below 25 years declines. That is, $\alpha_3 < 0$ (refer to equation (3.4.1)).

Hypothesis 4

Persons in the age group 25 – 50 years are much less likely to visit hospital and use hospital cares and services in contrast to persons of old age groups. The probability of using hospital cares and services for persons in
the age group 25 – 50 years declines. That is, \( \alpha_4 < 0 \) (refer to equation (3.4.1)).

**Hypothesis 5**

Persons in the age group above 50 years are much more likely to visit hospital (health care centers) and use health cares and services than the persons in other age groups. The probability of using hospital cares and services for persons above 50 years increases. That is, \( \alpha_5 > 0 \) (refer to equation (3.4.1)).

**Hypothesis 6**

The urban people are much more likely to visit hospital (health care centers) and use health cares and services than the rural people. The probability of using hospital cares and services for the residents in the urban areas is greater than that of the residents in the rural areas. That is, \( \alpha_6 > 0 \) (refer to equation (3.4.1)).

**Hypothesis 7**

Literacy is likely to have positive impact on the decision to visit health care centers and rate of utilization of health cares and services. The literate persons are much more likely to visit the health care centers and use health cares than the illiterate persons. The probability of using hospital care and services for educated persons is expected to increase compared to that of the non-educated persons. That is, \( \alpha_7 > 0 \) (refer to equation (3.4.1)).
Hypothesis 8

The frequency of illness is expected to affect the decision to visit health care centers directly. The probability of visiting hospital and using hospital services and cares of those who have high frequency of illness is likely to increase. That is, \( \alpha_8 > 0 \) (refer to equation (3.4.1)).

Hypothesis 9

Health conscious persons are much more likely to visit hospital (health care centers) and use health cares and services than the health unconscious persons. The probability of using hospital cares and services for health conscious persons is expected to increase. That is, \( \alpha_9 > 0 \) (refer to equation (3.4.1)).

Hypothesis A10

Income is likely to affect the decision to visit health care centers and the rate of utilization of health cares and services directly. The probability of visiting hospital and using hospital services and cares there will increase as the income increases. That is, \( \alpha_{a10} > 0 \) (refer to equation (3.4.1A)).

Hypothesis L10

Persons in the lower income group (up to Rs.36000 per annum) are much less likely to visit hospital and use hospital cares and services in contrast to persons of other income groups. The probability of using hospital cares and services for persons in the lower income group declines. That is, \( \alpha_{l10} < 0 \) (refer to equation (3.4.1B)).
Hypothesis M10

Persons in the middle income group (Rs.36000 – Rs.80000) are much more likely to visit hospital (health care centers) and use health cares and services than the persons in lower income groups. The probability of using hospital cares and services for persons in the middle income group increases. That is, $\alpha_{M10} > 0$ (refer to equation (3.4.B)).

Hypothesis U10

Persons in the upper income group (Rs.80000 and above) are much more likely to visit hospital (health care centers) and use health cares and services than the persons in other income groups. The probability of using hospital cares and services for persons in the upper income group increases. That is, $\alpha_{U10} > 0$ (refer to equation (3.4.B)).

Hypothesis 11

We cannot beforehand predict the impact of any particular community characteristics on the decision to visit health care centers and the rate of utilization of health cares and services unless the model is completely estimated. Since socio-economic and demographic and also cultural factors, associated with persons belonging to any particular community are very much important in taking some decision about the rate of utilization of health cares and services, this will enable us to make some hypotheses about the impact of religion and caste on the decision to visit health care centers.

We expect that persons who belong to the Hindu community are much more likely to visit health care centers and use health cares and services than
the persons who belong to non-Hindu community. Thus, the probability of using hospital care and services for the Hindus increases compared to that of the non-Hindus. That is, \( \alpha_{11} > 0 \) (refer to equation (3.4.1)).

**Hypothesis 12**

Persons who belong to the community of general caste are much more likely to visit health care centers and use health cares and services than the persons who belong to other non-general castes. Thus, the probability of using hospital cares and services for the general caste people increases compared to that of the non-general caste people. That is, \( \alpha_{12} > 0 \) (refer to equation (3.4.1)).

**Hypothesis 13**

The probability of using hospital cares and services for the people belonging to the Scheduled Caste, Scheduled Tribe and OBC is expected to be low compared to that of the general caste people. That is, \( \alpha_{13} < 0 \), \( \alpha_{14} < 0 \) and \( \alpha_{15} < 0 \) (refer to equation (3.4.1)).

3.5.2. **Hypothesis Relating to the Model for Out-of-Pocket Expenditure**

When the person visit health care centers and uses health cares and services, it is natural that the cost of treatment is to be borne. The cost of treatment made by a person who is protected by health insurance is expected to be smaller in quantity than the cost of treatment made by a person who is not protected by health insurance. Other than this fact of health insurance scheme, there are several other factors that would have also effect on the
amount of out-of-pocket expenditure for medical treatment. In this section let us set some hypotheses that we like to test empirically in the district of Birbhum on the basis of our sample collected during 2002 – 2003. The out-of-pocket payment for treatment has been modeled by using a log-linear function as specified in section 3.4.2.

**Hypothesis 1**

The health insurance scheme reduces the out-of-pocket payment for treatment. The out-of-pocket payment for persons protected by health insurance is smaller contrasted with the out-of-pocket payment for persons not protected by health insurance. That is, $\beta_1 < 0$ (refer to equation (3.4.2)).

**Hypothesis 2**

Males spend much more on their health care uses and services in contrast to females. The out-of-pocket payment for male is greater than that of females. That is, $\beta_2 > 0$ (refer to equation (3.4.2)).

**Hypothesis 3**

Persons in the young age group, that is, those who belong to the age group 1 with age below 25 years, spend less on health care. The out-of-pocket payment for the persons in this age group is smaller than that of persons in old age group. That is, $\beta_3 < 0$ (refer to equation (3.4.2)).

**Hypothesis 4**

Persons in the middle age group, that is, those who belong to the age group 2 with age in between 25 years to 50 years, also spend less on health
care. The out-of-pocket payment for the persons in this age group is smaller than that of persons in old age group. That is, \( \beta_4 < 0 \) (refer to equation (3.4.2)).

**Hypothesis 5**

Persons in the old age group, that is, those who belong to the age group 3 with age above 50 years, spend significantly high amount of money on health care. The out-of-pocket payment for the persons in this age group is higher than that of persons in other age groups. That is, \( \beta_5 > 0 \) (refer to equation (3.4.2)).

**Hypothesis 6**

The regional difference in the residence of the persons results in difference of cost of treatment. The cost of treatment in the urban area is smaller than that in the rural area. The urban people spend much less on health care than that by the rural people. The out-of-pocket payment for the persons in the urban area is smaller than that of persons in the rural area. That is, \( \beta_6 < 0 \) (refer to equation (3.4.2)).

**Hypothesis 7**

Literacy has got somewhat depressing impact on the health care expenditure. The out-of-pocket payment for the literate persons is expected to be smaller than that of persons who are not literate. That is, \( \beta_7 < 0 \) (refer to equation (3.4.2)).
Hypothesis 8

Frequency of illness is likely to increase health care expenditure. The out-of-pocket payment of the persons who frequently become ill will certainly be higher than the out-of-pocket payment of the persons who are not so. That is, $\beta_s > 0$ (refer to equation (3.4.2)).

Hypothesis 9

If the type of illness is complicated, the health care expenditure will increase. The out-of-pocket payment of the persons who are suffering from complicated disease will be naturally higher. That is, $\beta_s > 0$ (refer to equation (3.4.2)).

Hypothesis 10

Severity of illness will induce the cost of treatment to increase, since it requires too many days to stay in the health care centers. Therefore, the out-of-pocket payment for person attacked by severe illness is likely to increase. That is, $\beta_{i0} > 0$ (refer to equation (3.4.2)).

Hypothesis 11

Health consciousness is likely to reduce the cost on the health care expenditure. The out-of-pocket payment for the health conscious persons is expected to be smaller than that of persons who are not so. That is, $\beta_{i1} < 0$ (refer to equation (3.4.2)).
**Hypothesis A12**

Health care expenditure is likely to vary directly with the level of income. The out-of-pocket payment for treatment will rise as the level of income rises. That is, $\beta_{A12} > 0$ (refer to equation (3.4.2A)).

**Hypothesis L12**

The out-of-pocket payment on use of health cares and services for the persons who belong to the lower income group is expected to rise. That is, $\beta_{L12} > 0$ (refer to equation (3.4.2B)).

**Hypothesis M12**

For the persons/households who belong to the middle income group the out-of-pocket payment on use of health cares and services will increase. That is, $\beta_{M12} > 0$ (refer to equation (3.4.2B)).

**Hypothesis U12**

The out-of-pocket payment on use of health cares and services for the persons who belong to the upper income group will definitely rise. That is, $\beta_{U12} > 0$ (refer to equation (3.4.2B)).

**Hypothesis 13**

There is no definite relation between the community characteristic as religion and the health care expenditure. Thus, we write that $\beta_{13} \neq 0$ (refer to equation (3.4.2)).
**Hypothesis 14**

The health care expenditure of the persons or households who belong to the general caste is expected to increase. Therefore, the out-of-pocket payment of the general caste people will increase. That is, \( \beta_{14} > 0 \) (refer to equation (3.4.2)).

**Hypothesis 15**

For the persons who belong to the non-general caste, for example, Scheduled Caste, Scheduled Tribe and OBC, nothing can be confirmed about the direction of change in their health care expenditure. Therefore, we write that \( \beta_{15} \neq 0, \beta_{16} \neq 0 \) and \( \beta_{17} \neq 0 \).

3.5.3. Hypothesis Relating to the Model for Demand for Health Insurance

So far in the previous sections, the hypotheses relating to the rate of utilization of health cares and services and the corresponding out-of-pocket payment made by the persons or households have been presented. In the foregoing analysis, of all hypotheses the most important hypothesis of our interest is that which shows the impact of the health insurance on the decision to visit health care centers or rate of utilization of health cares and services and on the out-of-pocket payment of those who utilized health cares and services. Depending on the fact how health insurance affects the decision to visit health care centers or rate of utilization of health cares and services and the out-of-pocket payment, the households or persons will come forward to decide whether to demand for health insurance or not. Therefore,
let us now present a set of hypotheses that narrates the relation between the demand for health insurance and its several determinants in the district of Birbhum.

**Hypothesis 1**

Sex discrimination in favour of males and adversely against females increases the demand for health insurance in favour of males. Therefore, the males are much more likely to demand for health insurance. The probability of decision to purchase health insurance by the males is expected to increase. That is, $\gamma_1 > 0$ (refer to equation (3.4.3)).

**Hypothesis 2**

Persons who belong to the young age group, that is, the age group 1 with age below 25 years, are less likely to demand for health insurance. Therefore, the probability of decision to purchase health insurance by the persons in the age group 1 will decline. That is, $\gamma_2 < 0$ (refer to equation (3.4.3)).

**Hypothesis 3**

Persons who belong to the middle age group, that is, the age group 2 with age between 25 years to 50 years, are more likely to demand for health insurance. Therefore, the probability of decision to purchase health insurance by the persons in the age group 2 will increase. That is, $\gamma_3 > 0$ (refer to equation (3.4.3)).
Hypothesis 4

Persons who belong to the old age group, that is, the age group 3 with age above 50 years, are more likely to demand for health insurance. Therefore, the probability of decision to purchase health insurance by the persons in the age group 3 will increase. That is, $\gamma_4 > 0$ (refer to equation (3.4.3)).

Hypothesis 5

Male literacy tends to increase the demand for health insurance. Therefore, the probability of decision to purchase health insurance in such case will increase. That is, $\gamma_5 > 0$ (refer to equation (3.4.3)).

Hypothesis 6

Female literacy is expected also increase the demand for health insurance. Even we may think of some empowerment of literate women in the decision making of the family. This will definitely enable the women to exert their power and opinion that would affect the decision to purchase health insurance favourably. Therefore, the probability of decision to purchase health insurance in such case will increase. That is, $\gamma_6 > 0$ (refer to equation (3.4.3)).

Hypothesis 7

Large family may feel the strong need for health insurance, but economic capacity may not help them in realizing the need. The same idea may not be true for the small family. There may be some other logic behind the demand of small family for health insurance. However, family size may
have therefore positive or negative impact on the probability of decision to purchase health insurance. That is, $\gamma, \neq 0$ (refer to equation (3.4.3)).

**Hypothesis 8**

Worker population ratio and dependency ratio are two complimentary items affecting the demand for health insurance. Worker population ratio and the demand for health insurance are directly related. Therefore, the households with high worker population ratio are much more likely to demand for health insurance than the households with high dependency ratio. This implies that the probability of decision to purchase health insurance for the households with high worker population ratio will increase. That is, $\gamma_8 > 0$ (refer to equation (3.4.3)).

**Hypothesis 9**

Regional differences in the inhabitance of the households are important in affecting the demand for health insurance. We expect that the persons or the households in the urban area will demand for health insurance more than the persons or households in the rural area. Thus, the probability of decision to purchase health insurance of the urban people will increase. Hence, $\gamma_9 > 0$ (refer to equation (3.4.3)).

**Hypothesis 10**

Health consciousness affects the demand for health insurance favourably. The health conscious people will much more likely demand for health insurance than the health unconscious people. Thus, the probability of
decision to purchase health insurance for health conscious people will increase. That is, $\gamma_{10} > 0$ (refer to equation (3.4.3)).

**Hypothesis 11**

The greater the illness ratio the greater will be the demand for health insurance. This ensures that the households with high illness ratio are much more likely to demand for health insurance than the households with low illness ratio. Thus, the probability of the households with high illness ratio will increase. That is, $\gamma_{11} > 0$ (refer to equation (3.4.3)).

**Hypothesis 12**

If the households are properly aware of the health insurance schemes, they can take the right decision in the purchase of the scheme. This may increase the demand for health insurance. The households who are well informed of the terms and conditions of health insurance schemes (provided these schemes come to the benefits of the persons) will much more likely demand for health insurance that those who are not so. Therefore, the probability of the properly aware persons of the decision to purchase health insurance will increase. Thus, we expect $\gamma_{12} > 0$ (refer to equation (3.4.3)).

**Hypothesis 13**

The households with high annual health care cost per capita will feel strong need for health insurance. Such families are more likely to create the demand for health insurance. The probability of decision to purchase health insurance in such cases is expected to increase. That is, $\gamma_{13} > 0$ (refer to equation (3.4.3)).
Hypothesis 14

The households with high annual income per capita are strong enough to purchase health insurance. Such families are more likely to create the demand for health insurance. The probability of decision to purchase health insurance in such cases is expected to increase. That is, $\gamma_{14} > 0$ (refer to equation (3.4.3)).

Hypothesis 15

The higher the premium to be paid in relation to the benefit of the health insurance the lower will be the demand for health insurance. Thus, the premium benefit ratio and the probability of decision to purchase health insurance are inversely related. That is, $\gamma_{15} < 0$ (refer to equation (3.4.3)).

Hypothesis 16

Religion may not have any specific directional impact on the probability of decision to purchase health insurance. That is, $\gamma_{16} \neq 0$ (refer to equation (3.4.3)).

Hypothesis 17

Persons or households who belong to the general caste are more likely to demand for health insurance than the persons or households who belong to the non-general castes. Therefore, we expect that the probability of decision to purchase health insurance by the general caste people will increase. That is, $\gamma_{17} > 0$ (refer to equation (3.4.3)).
Hypothesis 18

Persons in the community of Scheduled Caste, Scheduled Tribe and OBC may not have any specific directional impact on the probability of decision to purchase health insurance. That is, \( \gamma_{18} \neq 0 \), \( \gamma_{19} \neq 0 \) and \( \gamma_{20} \neq 0 \) (refer to equation (3.4.3)).

In order to conclude this section, we would like to point out that after the specification of the econometric model and the corresponding hypotheses for different models relating to the issue of demand for health insurance, we will estimate the models and test the hypotheses using the sample observations collected for households in the district of Birbhum for the period 2002 – 2003. In the next section we therefore present the methodology of data collection.

3.6. The Case Study in Birbhum District

We have specified above three models that are to be estimated empirically and tested in the light of data in the district of Birbhum. With this end in view, we carried out a sample survey in the district of Birbhum for almost a year starting from mid July of 2003. The data was collected through personal interview method on the basis of some selected questions that include the information on the individual characteristics, household characteristics and the community characteristics. Since, the demand for health insurance is linked with the decision of the head of the family mainly,
we consciously incorporated some specific questions on the individual characteristics of the head of the households.

The district of Birbhum in West Bengal is not at all a prosperous district, agriculturally or industrially. It’s district of extreme whether in the degree of heat and cold with a tendency towards proneness to draught. Of the total population about 80% reside in the rural areas. Though poor, still agriculture is the principal occupation and source of income earning. Almost more than 65% of total population in the district of Birbhum is involved with agricultural activities in different status.

Paddy is the dominating crop of cultivation, followed by potato and mustard seeds. Sugar cane is also cultivated in some small plots of land. The cultivators are now considering vegetables for production. Though several efforts were taken consciously to diversify cropping patterns, it did not come out successful. Prices of agricultural crops are not fair to the original producers of the crops. The industrialization based on agricultural product has not been developed. Roads and communication are very poor. The benefits from agriculture therefore are not reaped by those who are involved in agriculture, but by the middlemen. Trade from this district to others therefore is not flourishing.

There is widespread poverty in the rural area of the district of Birbhum and this is notable among the rural people. Everyday a significant proportion of rural people rushes to urban areas simply to work as casual labour in the unorganized sector. Thus, there is also a huge scope of under
wage and exploitation of this section of people. The economic condition of the district is of so pitiable nature that it obviously has therefore made the population vulnerable to physical ailing. Malaria, diarrhea, dirig (some sort of fever), tuberculosis are the variety of illness the people are exposed to.

The health care situation is not satisfactory. Although, there are few primary health care centers run by the Government, the quality of medical cares and services is so horrible that people are very much averse to them. Medicines are not available. Doctors are not present when required. Number of patients is so high in relation to the available medical infrastructures. The facility of indoor treatment is almost nil in these rural health care centers. For a slightly complicated case of illness, the patients are advised to visit the district hospital, which is a far off from the rural palaces of the district of Birbhum. Due to road, transport and financial incapacity most of the patients die on its way. The plight of private health care centers is almost of same type. Not only that the cost of treatment in this private sector is exorbitantly high but also it is beyond the financial capacity of the common people (mainly rural people and casual labour) in the district of Birbhum. Access to health care is thus constrained by financial incapacity and inadequate health facilities accessible to the population of the district of Birbhum. Therefore, in most of the cases, in spite of high-risk disease the people are suffering from, they are simply left to their destiny. They sell even all their assets, sometimes are forced to sell their small plots of land for carrying out treatment and ultimately get pauperized landless labour.
Persons working in the organized sector get sometimes a partial reimbursement of their medical expenses. In contrast, the persons in the unorganized sectors are at sea. The situation is similar for the petty businessman, small and marginal farmers and self-employed persons. In the district of Birbhum, there is no such thing as community based health insurance scheme. Neither there is any network of medical team as found in rural Senegal (Tine, 2000) that can offer medicines to ailing persons at low cost, nor there is any such Government network that can arrange to send some doctors on regular basis to visit the rural places and serve the patients.

The issue of health insurance is therefore very much important to study. It can come to the help of common people providing financial assistance and benefits whenever necessary. As a matter of policy in the district of Birbhum health insurance is not at all compulsory, rather it is optional or voluntary. Anybody in exchange for some prepayment/premium can ensure some financial protection in future for his/her medical treatment.

3.6.1. Methodology of Data Collection

Because of the existence of formal health insurance market, it had been easier for us to conduct a survey and collect primary data through random sampling method. The data has been collected from different blocks in the district of Birbhum using random sampling method on multistage basis (method of three stage stratified random sampling). In the first stage of the sampling we select the block at random using simple random sampling method. Selection of village and town comes in the second stage of
sampling. In the final stage of sampling we choose the households in each village and town at random by the method of simple random sampling with replacement using the draw of random number. In order to project the complete picture from all sections and localities (rural and urban areas), we have divided the people of these blocks on the basis of area of residence, which fall in the developed and underdeveloped regions in the district of Birbhum. It is assumed that within each block, the population is homogeneous with respect to income, availability of health care facilities and other facilities. A block contains a number of villages. On the assumption of homogeneity, the respondents were selected randomly from each village and town and thus the complete sample of each block is formed.

We have considered ten blocks out of total nineteen blocks in the district of Birbhum and in each block two villages and two towns are selected. The choice of the village and the town in each block has been done on the basis the distance of these palaces from the private or public health care centers. One village or town, considered for data collection, is nearest to the health care center and the other village or town is farthest away from the health care center. Thus, in each of the forty rural (village) and urban (town) places roughly twenty households were selected at random. Thus, a sample of about 800 households has been considered for this study. The primary data for the period 2002 – 2003 was collected from different blocks, viz. Suri – I, Suri – II, Sainthia, Rampurhat, Murarai, Mayureswar, Bolepur, Dubrajpur, Labpur and Nanoor. It should be pointed out here that the district of Birbhum has been chosen by the purposive sampling method. The district of Birbhum, which is the residential place of this researcher, has been selected on personal judgments.
The number of the sample households according to main source of income is shown in table – 3.1.

Table – 3.1. Number of Households in the District of Birbhum.

<table>
<thead>
<tr>
<th>Category of Households</th>
<th>Rural</th>
<th></th>
<th></th>
<th>Urban</th>
<th></th>
<th>District</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-insured</td>
<td>Insured</td>
<td>Total</td>
<td>Non-insured</td>
<td>Insured</td>
<td>Total</td>
</tr>
<tr>
<td>Number of Households</td>
<td>217</td>
<td>163</td>
<td>380</td>
<td>152</td>
<td>252</td>
<td>404</td>
</tr>
<tr>
<td>Service</td>
<td>88</td>
<td>84</td>
<td>172</td>
<td>51</td>
<td>133</td>
<td>184</td>
</tr>
<tr>
<td>Cultivation</td>
<td>63</td>
<td>40</td>
<td>103</td>
<td>58*</td>
<td>71*</td>
<td>129*</td>
</tr>
<tr>
<td>Casual Labour</td>
<td>26</td>
<td>10</td>
<td>36</td>
<td>15</td>
<td>18</td>
<td>33</td>
</tr>
<tr>
<td>Business</td>
<td>40</td>
<td>29</td>
<td>69</td>
<td>28</td>
<td>30</td>
<td>58</td>
</tr>
</tbody>
</table>

* Number of persons engaged in agriculture but residing in the urban area of the district.

The consciousness of the respondents regarding health care and its problem is an important factor of collection of data. If the respondents are aware of the health care problem, it is expected that their repercussions will be captured/reflected in the response. Before the collection of data from each interviewee, it is necessary to portray the health care problem.

Recent development in the health care sectors, owing to new invention and modern equipments have resulted in a rapid increase in health care expenditure of the people and imposing tremendous pressure on the people thereby. Though the public health care sectors are providing the health cares and services almost free of cost or at a very low and subsidized cost, it could not mitigate the demand for health care of the people. So the insufficient supply of health cares and services by the public sectors has created financial pressure on the people due to treatment in the private sector health care centers. Further, the inadequate and poor quality of health cares
and services provided by the public sector health care centers has led the people in the district of Birbhum to depend on the alternative private sector health care centers, which can play a significant role. The health care realities and the problems of the privately run health care centers also provide crucial information in this study. At the backdrop of this information, we prepared our questions. Had this not been done, the responses of the interviewees could not focus properly on the coming crisis. Naturally in the absence of any future perspective in mind, there could be hardly any decision to purchase health insurance for improved health cares and services. Sometimes the people were misusing/abusing the public sector health cares and services for their minor or negligible diseases and the Government continued providing subsidies for health cares and services. This has been the story till recently, while the Governments have started changing their approach of subsidizing health cares and services to the people. So for the efficient health cares and services, it is equally important to understand the reactions of the people at the backdrop of the above mentioned information.

While collecting data it was apparent that a prior discussion of the coming crisis (at places it is going on) in the health cares and services in the district of Birbhum would enable the respondents to think rationally and put forward for better health cares and services. As the major focus of the study deals with the demand for health insurance, it is important to have information on its various determinants. Broadly speaking these factors are: premium for health insurance, income of the person concerned, cost of health cares, price of medicines, possibility of tax exemption, educational
background, preferences of health care, degree of health consciousness, worker population ratio or the dependency ratio in family etc. We divided categorically all the relevant information into individual, household and community characteristics. All these characteristics are merely composed of economic and socio-economic demographic features of the respondents. Above all the economic factors are of prime importance and income and prepayment for the health insurance scheme constitute the most important components. So an integral part of the efficient and effective health cares and services is to persuade the people to take the decision to purchase health insurance. However, in the urban area where the quality of health cares and services is good, the relevant question is whether the persons or households are to take the decision to purchase health insurance to retain the facilities at the present level. This is because of the fact that people are facing severe financial crunch to uphold the present set of health cares and services. Further, the existing price structure of health cares and services is neither market determined nor does it portray the possible consequences of the demand – supply mismatch.

The principal question is: *Is it possible for the public sector to take initiatives of future demand for health cares and services? If the answer is ‘no’, what would be the decision of the people about the purchase of health insurance?* This question will be answered in a better way if we look at the estimation of the demand for health insurance in the district of Birbhum.
3.6.2. Diagnostic Check for Sample Size

In many empirical applications the problem of disproportionate sampling is very much common. It is often found that the number of observations in one of the groups is much smaller than the number in the other group. In the case of present study the problem of such disproportionate sampling could have appeared if the number of households in the urban area were significantly greater than the number of households in the rural area. This would have led some urban bias and consequently the sample chosen may not be a representative sample. We have tried to solve this problem in this study.

In our study we have always tried to decide the sample size of the households in such a manner that we can escape the bias either to the rural areas or to the urban areas in the district of Birbhum. Therefore, we had to adjust often the number of households from rural and urban areas so that the sample proportions of households in the entire population from the rural areas and from the urban areas in the district of Birbhum do not differ significantly. We have decided finally the sample size by the test of significance of the difference of sample proportion of households in the rural area and in the urban area. We have to carry out a test based on standard normal statistic \( Z = \left( \frac{(p_1 - p_2)}{\sqrt{pq(l/n_1 + l/n_2)}} \right) \) which is \( N(0,1) \). On the basis of this test that is designed to test the significance of the difference of sample proportion of households in the rural area and in the urban area, the sample size was finalized.
3.7. Conclusion

Households or persons are classified, first of all, into two groups, one group protected by health insurance and the other group not protected by health insurance. The fact that one group of persons is protected by health insurance and the other group is not protected by health insurance may be explained by different individual, household and community characteristics of the sample population. We shall try to explain this fact in the light of our data collected in the district of Birbhum on different socio-economic and demographic aspects. We would carry out a test of significance of difference between the mean of different explanatory variables for the group of persons protected by health insurance and the corresponding mean for the group of persons not protected by health insurance. If this difference in means for any particular variable appears statistically significant, we shall conclude that the variable in question may be the factor affecting the decision to purchase health insurance.

In an alternative fashion the factors of demand for health insurance may be gauged. This is simply the Pearsonian product moment correlation coefficient between the decision to purchase health insurance and the several explanatory variables. Then one should carry out the test of correlation coefficient using t-statistic. If the product moment correlation coefficient between the decision to purchase health insurance and any variable, say level of income, is statistically significant, we should conclude income is significant variable in affecting the decision to purchase health insurance.
In this chapter we have presented the analytical framework of the models that would be used in the study of demand for health insurance. After the definitions and specifications of the actual variables affecting the decision to visit health care centers/ rate of utilization of health cares and services, out-of-pocket payment for health cares and services and the decision to purchase health insurance, we have specified the actual models in line with analytical framework. As per our objectives of this study, several hypotheses relating to the several models of demand for health insurance have been clearly formulated. In order to estimate and test these hypotheses the data collected in the district of Birbhum for the period 2002 – 2003 will be used. The methodology of collecting data in the district of Birbhum for the period 2002 – 2003 has been also presented in this chapter. Now, using the collected sample observations, we would estimate the models of demand for health insurance and test the hypotheses we have set and it is going to be presented in the next chapter (chapter four).