Chapter – II

Materials and Methods
Chapter – II
Materials and Methods

2.1 Introduction

The present chapter deals with the methodological approaches adopted in the present investigation. The basic terminologies, analytical methods and their mathematical aspects form the core part of present study.

Terminologies

**Birth weight:** Birth weight is the first weight of the foetus or newborn obtained after birth. For live births, birth weight should preferably be measured within the first hour of life, before the occurrence of significant postnatal weight loss (UNICEF, 2004).

**Low Birth Weight (LBW):** The prevalence of low birth weight is defined as the percentage of live births that weigh less than 2500 grams out of the total number of live births during the same time period. The low birth weight prevalence rate is calculated as (UNICEF, 2004).

\[
\text{Number of live born babies with less than 2500 grams birth weight} \times 100
\]
\[
\text{Number of live births}
\]

2.2 Materials

The third round of National Family Health Survey (NFHS-3) data was used under this research work (2005-2006). National Family Health Survey (NFHS-3) covered all 29 states of India, which comprise more than 99 percent of India’s population. Birth weight was recorded in the questionnaire for births occurring during the five years preceding the survey from available reports or reported by the mother’s (IIPS, 2007). The following questions were asked during the time of survey:

(i) - When (NAME) was born, was he/she very large, larger than average, average, smaller than average, or very small?
(ii) - Was (NAME) weighed at birth?
(iii) - (IF YES) How much did (NAME) weigh?
Birth weight is an important indicator of a child’s vulnerability to the risk of childhood illness and chances of survival. In the absence of birth weight, a mother’s subjective assessment of the size of the baby at birth is a useful proxy for birth weight. Children whose birth weight is less than 2.5 kilogram’s (2500 grams), or children reported to be ‘very small’ or ‘smaller than average’ are considered to have a higher than average risk of early childhood death. Since birth weight may not be known for many babies, the mother’s estimate of the baby’s size at birth was obtained for all births (IIPS, 2007). The National Family Health Survey (NFHS-3) data provides a comprehensive picture of population and health conditions in India. The data on the LBW are available in public domain.

2.2 Methods

2.3.1 Variables

Birth weight was reported from mother memory recall or health card. As per recall mothers produced approximate birth weight or perception of birth size. In this study birth weight was categorized from original response variable to dummy variable. Birth weight was used as dependent variable. It was categorized in two groups as: “1” if Low Birth Weight and “0” otherwise. Low Birth Weight means children weighing less than 2500 grams. Otherwise means normal birth weight who weighed 2500 grams or more. The reported birth weights were categorized and all other response were considered under not reported category of birth weight. Birth weight reporting was also used as dependent variable and it was categories in two categories. Predictor variables are also discussed in detail. Definition and classification of variables are mentioned in following table.

Table: 2.1, shows the description of considered variables under the study.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Response categories</th>
<th>Percent</th>
<th>Description of variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth weight Reporting</td>
<td>Reported</td>
<td>34.10</td>
<td>Either health card or memory recall</td>
</tr>
<tr>
<td></td>
<td>Not-reported</td>
<td>65.90</td>
<td>Not-weighed + special answer = Not-reported</td>
</tr>
<tr>
<td>Birth Weight</td>
<td>Low</td>
<td>21.50</td>
<td>Low birth weight &lt;2500g</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>78.50</td>
<td>Normal birth weight ≥ 2500g</td>
</tr>
<tr>
<td>Body Mass</td>
<td>&lt; 18.5 kg/m2</td>
<td>38.84</td>
<td>The body mass index (BMI) is the</td>
</tr>
<tr>
<td>Variable</td>
<td>Value</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Index (BMI)</td>
<td></td>
<td><strong>18.5 to &lt;25 kg/m²</strong></td>
<td>ratio of the weight in kilograms to the square of the height in meters (kg/m²). Underweight is Less than 18.5 kg/m²; Normal weight =18.5 to &lt;25 kg/m²; Overweight = 25 and above kg/m²</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>25 &amp; above kg/m²</strong></td>
<td></td>
</tr>
<tr>
<td>Mother’s Anemia Status/level</td>
<td></td>
<td>Anemic Mother’s</td>
<td>Iron deficiency anemia characterized by low level of hemoglobin in the blood. If the hemoglobin level is 10.9 g/dl of pregnant women= No anemia and if Hb&lt;10.9 g/dl= anemia. We have combined mild, moderate and severe anemia into one category.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not anemic Mother’s</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not anemic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No ANC Visit</td>
<td>Antenatal care during pregnancy created as dummy variable in three groups from original variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-2 ANC Visit</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3+ ANC Visit</td>
<td></td>
</tr>
<tr>
<td>Wealth index</td>
<td></td>
<td>Poor</td>
<td>Wealth index created as dummy variable in three groups from original variable. It is as Poorer + poorest = Poor, Richer + Richest = High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Residence</td>
<td></td>
<td>Rural</td>
<td>Place of residence is used original</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urban</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>Uneducated</td>
<td>Education is also used as dummy variable. It is categories into three groups from original variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Educated but &lt; Secondary</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary &amp; Above</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>15-24 age group</td>
<td>Age of the mother at the time of survey in completed years but it is categories in 3 age groups from single years age variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25-34 age group</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>35-49 age group</td>
<td></td>
</tr>
<tr>
<td>Birth order</td>
<td></td>
<td>First order</td>
<td>Birth order variable is categories only five orders for analysis purpose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second order</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Third order</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fourth order</td>
<td></td>
</tr>
</tbody>
</table>
Table: 2.1, demonstrates detailed description of created dummy variable from the original variables, used in the analysis. The selections of variables were based on their availability in the data set. The proximate factors identified for cause of low birth weight like, mother’s factor, child factor and socio-demographic factors were considered in analysis. Mothers body mass index (BMI) was divided into three categories, underweight (BMI < 18.5), normal weight (18.5 to <25) and overweight (25 and above) and it was considered as a nutritional indicator. BMI was a reliable indicator of chronic energy deficiency (Shetty PS et al., 1994). Body mass index was measured during the preconception period to examine the effect on LBW. The data distribution suggested that 23% of mothers had never visited for ANC check-up.

2.3.2 Accuracy in birth weight data- Diagnostic Test (Birth weight Vs. Birth size)
The assessment of any phenomena relies on comparative decisions based on actual facts (truth). The ‘truth’ in the present context refers to whether the disease is present or absent in an individual. In terms of birth weight data, truth is considered as in terms of presence or absence of numerical birth weight several. Bio-statistical approaches were considered for analysis of the accuracy of the birth weight data. Data regarding birth weight during the survey was reported through mother memory recall (numerical weight or mother’s perception on birth size) and health card.

2.3.3 2 X 2 contingency table for birth weight and birth size

It is well known fact that there limitations in the accuracy, or which indices, independent of the prevalence of the condition and thereby separating the various types of right and wrong conclusions, are desired. The possible situations in any diagnostic test, when the outcome is binary (Yes/ No), can be summarized in a 2 X 2 table as shown below. With the help of this 2 X 2 contingency table, we can easily understand the various steps of the analysis concepts of birth weight and birth size.

Table 2.2: 2 X 2 Contingency table for birth weight and birth size

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th>Birth weight reported in terms of birth size</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (Small Size)</td>
<td>Normal (Large Size)</td>
<td>Total</td>
</tr>
<tr>
<td>Numerical birth weight reported</td>
<td>Low Birth Weight (&lt; 2500 grams)</td>
<td>a (TP)</td>
<td>b (FN)</td>
</tr>
<tr>
<td></td>
<td>Normal Birth Weight (&gt; 2500 grams)</td>
<td>C (FP)</td>
<td>d (TN)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
</tr>
</tbody>
</table>

As we can see from the above table, the test can lead to a conclusion that True positive (TP): Proportion of children having birth weight below 2500 grams and reported as of small size. It is denoted by ‘a’. False negative (FN): Proportion of children having birth weight below 2500 grams but reported as of large size. It is denoted by ‘b’. True negative (TN): Proportion of children having birth weight above 2500 grams and also reported as of large size. It is denoted by ‘d’. False positive (FP):
Proportion of children having birth weight above 2500 grams but reported as of small size. It is denoted by ‘c’.

2.3.3.1 Sensitivity: The strength of birth size to correctly identify children with low birth weight (LBW- small size) as true under recorded birth weight. It is calculated as:

\[
\text{Sensitivity} = \frac{a}{a + b} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
\]

2.3.3.2 Specificity: The strength of birth size to correctly identify children with normal birth weight as true under recorded birth weight. It is calculated as:

\[
\text{Specificity} = \frac{d}{c + d} = \frac{\text{True Negative}}{\text{True Negatives} + \text{False Positive}}
\]

2.3.3.3 Positive predictive value (PPV): The proportion of children of small size who remained low birth weight as per recorded birth weight. It is calculated as:

\[
\text{PPV} = \frac{a}{a + c} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}
\]

2.3.3.4 Negative predictive value (NPV): The proportion of children of normal size who remained normal as per recorded birth weight. It is calculated as:

\[
\text{NPV} = \frac{d}{b + d} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}}
\]

2.4 Kappa Statistic

Cohen’s kappa statistic (after Jacob Cohen, 1960) is a measure of inter–rater agreement for categorical variables. The two raters are considered as two tests
independently. It is generally thought to be a more robust measure than simple percent agreement. The calculation in Kappa statistic takes into account the agreement occurring by chance. The $\kappa$ statistic can be calculated from a 2 x 2 or higher order symmetrical (i.e. same number of rows and columns) contingency table. The calculation of $\kappa$ is relatively simple and can be done by hand.

Kappa takes value between 0 and 1. If the raters or tests are in complete agreement then $\kappa = 1$. If there is no agreement among the raters or tests at all (other than what would be expected by chance) than $\kappa = 0$. Intermediate values are interpreted as follows though the boundaries depicted are not universally accepted. The number of categories and subjects will affect the magnitude of the value. The kappa tends to be higher when there are fewer categories.

### 2.4.1 Interpretation of $\kappa$

- $0.00 – 0.20$ Poor agreement
- $0.21 – 0.40$ Fair agreement
- $0.41 – 0.60$ Moderate agreement
- $0.61 – 0.80$ Substantial agreement
- $0.81 – 1.00$ Almost perfect agreement

Note that Cohen’s kappa measures agreement between two raters or tests only. Fleiss’ kappa (after Joseph L. Fleiss, 1981) statistic has been used for a similar measure of agreement in categorical rating when there are more than two raters (Avijit Hazra, 2013).

Mathematically, kappa statistic is defined and calculated as (Sundaram et.al, 2015)

$$\text{Kappa} = \left\{ \frac{\text{Observed Agreement (O)} - \text{Expected Agreement (C)}}{1 - \text{Expected Agreement (1- C)}} \right\}$$

$$O = \left\{ \frac{(a + d)}{n} \right\}$$

And,

$$C = \left\{ \left[ \frac{(a + c)}{n}(a + b)/n \right] + \left[ \frac{(b + d)}{n}(c + d)/n \right] \right\}$$
Where, \( O \) = Observed proportion of Agreement, \( C \) = Proportion of chance Agreement, 
\( a \) = True positive value, \( b \) = False negative value, \( c \) = False positive value, 
\( d \) = True negative value and,

\[
 n = a + b + c + d
\]

Theoretically, ‘\( n \)’ was total number of birth weight reported, which was presented in the data set in 2 X 2 contingency table. Kappa coefficient is considered to represent excellent agreement if the value is greater than 0.75. Moderate agreement considered if the value is lie between 0.40 to 0.75. Poor agreement if the value is less than 0.40 (Gayle HD et al., 1988).

### 2.5 Chi-square (\( \chi^2 \)) test

The values of \( \chi^2 \) are also calculated to check the association between birth weight reporting and different predictors. As usual the \( \chi^2 \) can be calculated as:

\[
\chi^2 = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i}
\]

Where \( e_i \)'s are the expected frequencies and \( O_i \)'s the observed frequencies.

It is to be mentioned that the value of \( \chi^2 \) can be considered to be an indicator of departure from expectation (expectation on the assumption of birth weight). However, the value of \( \chi^2 \) can be affected by the sample size (\( N \)). To overcome this difficulty the coefficient of contingency (as suggested by Karl Pearson) can be calculated:

\[
C = \sqrt{\frac{\chi^2}{\chi^2 + N}}
\]

Obviously larger the value of \( C \), larger is departure from expectation.

### 2.6 Logistic Regression Model

To describe the relationship between a quantitative response variable (dependent variable) and one or more explanatory variables generally simple or
multiple linear regression analysis is carried out. However, when response variable becomes dichotomous in nature, linear regression becomes inappropriate. For this condition, logistic regression methods are appropriate methods. Out of various methods available for analysis of such data, logistic regression analysis is preferred mainly because it involves mathematically flexible and easily used function and also leads to an epidemiological meaningful interpretation. Wider acceptability to logistic model was achieved when Truett, Cornfield and Kannel (1967) provided the multivariate analysis of Farmingham heart study. After this work, it has become the standard method for regression analysis of dichotomous data in many fields, especially in the field of epidemiological study.

The logistic regression model that specifies the probability of a birth weight depending on a set of n-explanatory variables \( X'=(X_1,X_2,..................,X_n) \) may be defined as (Kleinabum 1988; Hosmer and Lemeshow 1991):

\[
\lambda(X) = \frac{\exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_n X_n)}{1 + \exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_n X_n)}
\]

\( \lambda(X) \) Denotes the conditional probability that the low birth weight is present \( P(Y=1 \mid X) \), \( \theta 's \) are the parameters representing the effects of the \( X_n 's \) on the risk (probability) of low birth weight.

2.6.1 Maximum Likelihood Estimation

The general method of estimation that leads to the least square functions under the linear regression models (when the error terms are normally distributed) is called maximum likelihood. Method of maximum likelihood yields values for the unknown parameters that maximize the probability of obtaining the observed set of data.

If the response variable \( Y \) is coded as zero (Normal birth weight - NBW) and one (Low birth weight – LBW), then the expression for \( \lambda(X) \) in equation (1) provides (for arbitrary value of \( \theta '=(\theta_0,\theta_1) \)) the conditional probability that \( Y \) is equal to 1 given \( X \). This is written as \( P(Y=1 \mid X) \). This signifies that the quantity \( 1-\lambda(X) \) gives the conditional probability that \( Y \) is equal to zero given \( X \), \( P(Y=0 \mid X) \). Therefore for those pairs \( (X_i,Y_i) \), where \( Y_i=1 \) the contribution to the likelihood functions is \( \lambda(X_i) \), and for those pairs where \( Y_i=0 \) the contribution to likelihood function is \( 1-\lambda(X_i) \), where \( \lambda(X_i) \) denotes the value of \( \lambda(X) \) computed at the particular value of \( X \), say \( X_i \).
More explicit way to express the contribution of likelihood function for the pair \((X_i, Y_i)\) is as:

\[
\varphi(X_i) = \lambda(X_i)^{y_i} [1 - \lambda(X_i)]^{1-y_i} \quad \text{…………………………(2)}
\]

As observations are assumed to be independent, the likelihood function is obtained as the product of the terms given in expression (2) that is given in the following manner:

\[
\ell(\theta) = \prod_{i=1}^{n} \varphi(X_i) \quad \text{…………………………………………………………..(3)}
\]

According to the principle of maximum likelihood method, we use the estimate of \(\theta\) that maximizes the expression (3). Taking its logarithms both side of the equation yields log likelihood function as:

\[
L(\theta) = ln[\ell(\theta)] = \sum_{i=1}^{n} [y_i \ln(\lambda(X_i)) + (1 - y_i) \ln(1 - \lambda(X_i))] \quad \text{……………(4)}
\]

To find the value of \(\theta\) that maximizes \(L(\theta)\), the partial differentiations of expression (4) are carried out with respect to \(\theta_0\) and \(\theta_1\) respectively. The resulting expressions are set equal to zero, yielding likelihood equations:

\[
\sum_{i=1}^{n} [y_i - \lambda(X_i)] = 0 \quad \text{……(5)}
\]

and,

\[
\sum_{i=1}^{n} X_{ij} \cdot [y_i - \lambda(X_i)] = 0 \quad \text{……(6)}
\]

These expressions are nonlinear in the parameters. Therefore, they require special methods for solutions that are iterative in nature. Most of the standard statistical software possesses special programs for this analysis generally as a logistic regression model.

2.6.2 Significance of a Variable
The significance of the variables in the model can be assessed by comparing the models with and without the variable, which is based upon log likelihood function given by equation 4. Observed value of the response variable may also be considered as predicted value from the saturated model. The saturated model is the model containing as many parameters as desired data points. The comparison of observed to predict values using the likelihood functions rest on expression as:

\[ D = -2\ln \left( \frac{\text{Likelihood of the Current Model}}{\text{Likelihood of the Saturated model}} \right) \]  

(7)

The quantity inside the large bracket in expression (7) is termed as likelihood ratio. Minus twice its log likelihood is done to achieve known distribution useful for hypothesis testing. D in the equation (7) is termed as deviance. For assessing the significance of an independent variable, the value of D with and without the independent variable is compared. The change in D is:

\[ G = D \text{ (For Model with the variable)} - D \text{ (For Model without the variable)} \]

Since the likelihood of the saturated model is common to both of the values of D, G can be expressed as:

\[ G = -2\ln \left( \frac{\text{Likelihood without the Variable}}{\text{Likelihood with the Variable}} \right) \]  

(8)

Under the null hypothesis that \( \theta \) is equal to zero, the statistic G follows a Chi-Square distribution with 1 degree of freedom.

2.6.3 Logit Transformation:

For proper interpretation of the estimated coefficients, expression given in equation 1 may be transformed as:

\[ g(X) = \ln \left( \frac{\lambda(X)}{1-\lambda(X)} \right) = \left( \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots + \theta_p X_p \right) \]  

(9)

Equation (9) is termed as logit transformation. The importance of this transformation is that \( g(X) \) possesses many of the desirable properties of a linear regression model. The logit, \( g(X) \), is linear in its parameters and may also range from \(-\infty\) to \(+\infty\), depending upon range of values of X. Constructing the confidence interval
of the coefficients, inferences are usually based upon the sampling distribution of estimated coefficient which tends to follow a normal distribution even for smaller samples. Accordingly, 100 x (1-α)% Confidence Interval (C.I.) estimate for the odds ratio is obtained by calculating the endpoints of the coefficient \( \theta \), and then exponentiation in following manner:

\[
exp\left[ \theta^* \pm Z_{1-\alpha/2} \times SE(\theta^*) \right] \quad \text{.........(10)}
\]

2.6.4 Interpretation of Coefficients

Like in case of linear regression, the estimated regression coefficients represent the slope or rate of change in log logit function of the dependent variable as a result of per unit change in the independent variable. In the logistic model, \( \theta \) represents change in the logit for a one unit change in the covariates, \( X \), i.e., \( g(X+1) - g(X) \); where \( g(X) \) is the logit transformation defined by the expression (9). Understanding the meaning of difference between these two logits is important for adequate interpretation of the estimated coefficients. In case of consideration of only one dichotomous independent variable that is coded as 0 or 1, the values of the regression model may be expressed as described below in tabular form.

<table>
<thead>
<tr>
<th>Outcome variable (Y)</th>
<th>Independent variable (X)</th>
<th>X=1</th>
<th>X=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = 1 )</td>
<td>( \lambda(1) = \frac{exp(\theta_0 + \theta_1)}{1 + exp(\theta_0 + \theta_1)} )</td>
<td>( \lambda(1) = \frac{exp(\theta_0)}{1 + exp(\theta_0)} )</td>
<td></td>
</tr>
<tr>
<td>( Y = 0 )</td>
<td>( 1 - \lambda(1) = \frac{1}{1 + exp(\theta_0 + \theta_1)} )</td>
<td>( 1 - \lambda(1) = \frac{1}{1 + exp(\theta_0)} )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

Odds of the outcome (Low birth weight) being present among individuals with \( X=1 \) is defined as \( \lambda(1)/[1 - \lambda(1)] \). On the similar pattern odds of the outcome being present among individuals with \( X=0 \) is defined as \( \lambda(0)/[1 - \lambda(0)] \).

Therefore log of the odds, called logit, is defined as:

\[
g(1) = \ln \left[ \frac{\lambda(1)}{1 - \lambda(1)} \right] \quad \text{.................(11)}
\]
And, 
\[ g(0) = \ln \left[ \frac{\lambda(0)}{1-\lambda(0)} \right] \] ..........................(12)

The odds ratio, denoted as \( \vartheta \), is defined as the ratio of the odds for \( X=1 \) to the odds for \( X=0 \):
\[ \vartheta = \frac{\lambda(1)/1-\lambda(1)}{\lambda(0)/1-\lambda(0)} \] ..........................(13)

The log of the odds ratio, termed as log-odds ratio or log-odds is given by the equation:
\[ \ln(\vartheta) = \ln \left[ \frac{\lambda(1)/1-\lambda(1)}{\lambda(0)/1-\lambda(0)} \right] = g(1) - g(0) \] ............(14)

Now, using the expressions for the logistic regression model as shown in above tabular form, the odds ratio is:
\[ \vartheta = \left( \frac{e^{\theta_0+\theta_1}/1+e^{\theta_0+\theta_1}}{e^{\theta_0}/1+e^{\theta_0}} \right)^{\frac{1}{1+e^{\theta_0}}} = \frac{e^{\theta_0+\theta_1}}{e^{\theta_0}} = e^{\theta_1} \] .............(15)

Hence under the logistic regression analysis with a dichotomous independent variable,
\[ \vartheta = e^{\theta_1} \]
and the logit difference, or log odds, is
\[ \ln(\vartheta) = \ln(e^{\theta_1}) = \theta_1 \]

Therefore, the odds ratio approximates how much more likely (or unlikely) it is for the low birth weight to be present among those categories with \( X=1 \) than those among with \( X=0 \).

2.6.5 Univariate logistic regression model

To begin with, univariate logistic regression model was carried out taking single variables at a time and dependent variables as LBW/NBW. This analysis was carried out using one of the related modules in the statistical package “SPSS (16.0 version)”. The output of the results were obtained in the form of regression coefficient, standard error of estimated regression coefficient, risk ratio/ odds ratio
(now onward referred as odds ratio) and its 95% Confidence interval. As obvious, through this analysis, the estimated odds ratio will provide adjusted odds ratio/ crude odds ratio. The resulted so obtained in relation to dependent variables are listed in tabular form in the corresponding chapters.

2.6.6 Multivariable Logistic Regression Models

After finalization of the potential covariates by regression analysis multivariable logistic regression models were obtained corresponding to dependent variables as: Birth weight – LBW/NBW and Birth weight reporting – Reported/Not reported. Total of four models were developed in chapter-V. The epidemiological models were obtained by applying traditional logistic regression analysis in which the corresponding dependent variables and all the potential covariates were taken at a time. For fourth model in chapter-V, the results were obtained in the form of covariates odds ratio and 95% confidence interval.

2.7 Ratio Methods

The prevalence of low birth weight and its change due to heaping problem in collected data on birth weight the following techniques, are adjusted by Ratio Methods. This technique was proposed by (Channon et.al, 2011). ‘Ratio Method’, is defined as, the percentage apportioned varies determined by obtaining the total number of babies weighing 2,000–3,000 grams, excluding those weighing exactly 2,500 grams. The percentage of infants weighing 2,000–2,499 grams out of the total weighing 2,000–3,000 grams (excluding those weighing exactly 2,500 grams) is calculated. Assuming that the distribution of birth weights between 2,000 and 3,000 grams is linear, this percentage will be then used to reclassify the same percentage of those infants weighing exactly 2,500 grams into the LBW category (Channon et.al, 2011).

Mathematical calculation of the ratio method was considered under the revised estimate of the low birth weight and is also termed as adjusted LBW. The adjusted LBW obtained on the availability of birth weight data in the data set. The following steps demonstrate the calculation under ‘Ratio Method’.

Birth weight group (2000 – 2499 grams)

\[ = n_1 \text{ (frequency)} = 2835 \]
Birth weight group (2501 – 3000 grams)  

= n₂ (frequency) = 6324

Birth weight on exact weight 2500 grams  

= n₃ (frequency) = 3589

As per, definition low birth weight is <2500 grams  

= n₄ (frequency) = 4146

Total number of birth weight reported in the study data  

= n (frequency) = 19250

Proportion p₁ calculated as  

\[ p₁ = \frac{n₁}{n₁ + n₂} \]

\[ = \frac{2835}{2835 + 6324} \]

\[ = \frac{2835}{9156} = 0.31 \]

Proportion p₂ calculated as  

\[ p₂ = n₃ \times p₁ = 3589 \times 0.31 = 1111 \]

Adjusted low birth weight calculated as  

\[ = \frac{n₄ + p₂}{n} = \frac{4146 + 1111}{19250} \]

\[ = \frac{5257}{19250} = 27.31\% \]
In India, an unadjusted low birth weight was 21.5% and adjusted low birth weight was found 27.31%. It shows the heaping of birth weight on exact weight 2500 grams, which thereby affect the prevalence of LBW in India.

2.8 Epidemiological Logistic Regression Models

Model-I:

An univariate logistic regression model has been developed wherein birth weight category (LBW=1 and NBW=0) was used as outcome variable and nutritional factor body mass index (BMI = $x_1$) considered as predictor variable. Equation of this epidemiological model is:

$$\log \left[ \frac{p}{1-p} \right] = b_0 + b_1 x_1,$$

where $b_1$ was the regression coefficients and $\log \left[ \frac{p}{1-p} \right]$ is called log odds or logit of the event.

Model-II:

A bivariate logistic regression model has been developed wherein birth weight category (LBW=1 and NBW=0) was used as outcome variable and BMI = $x_1$ & Anemia = $x_2$ considered as predictor variable. Equation of the epidemiological model is:

$$\log \left[ \frac{p}{1-p} \right] = b_0 + b_1 x_1 + b_2 x_2$$

where $b_1$, $b_2$, bivariate regression coefficients and $\log \left[ \frac{p}{1-p} \right]$ is called log odds or logit of the event.

Model-III:

In model 3, we used three explanatory variables. In terms of three maternal factors trivariate logistic regression model has developed wherein birth weight category (LBW=1 and NBW=0) was used as outcome variable and BMI = $x_1$, Anemia = $x_2$ & Antenatal care = $x_3$ considered as predictor variable. Equation of the epidemiological model is:

$$\log \left[ \frac{p}{1-p} \right] = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3,$$

where $b_1$, $b_2$, $b_3$ were the logistic regression coefficients and $\log \left[ \frac{p}{1-p} \right]$ is called log odds or logit of the event.

Model-IV:

Here a multivariable logistic regression model has been developed wherein birth weight category (LBW=1 and NBW=0) was used as outcome variable and $x_1$
(Place of residence), $x_2$ (Age), $x_3$ (Wealth index), $x_4$ (Education), $x_5$ (Religion), $x_6$ (Caste), $x_7$ (Anaemia level), $x_8$ (Body mass index), $x_9$ (Antenatal care) and $x_{10}$ (Birth order) were considered as predictor variables.

Equation of the epidemiological model is:

$$\log \left[ \frac{p}{1-p} \right] = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + b_6 X_6 + b_7 X_7 + b_8 X_8 + b_9 X_9 + b_{10} X_{10},$$

where $b_1$, $b_2$, $b_3$, ..., $b_{10}$ are the logistic regression coefficients and $\log \left[ \frac{p}{1-p} \right]$ is called log odds or logit of the event.

### 2.9 Statistical Packages:

The complete analysis under the present study was accomplished by utilizing various licenses statistical packages, SPSS (16.0) and STATA (11.0) either available in the department of Statistics, Gauhati university, Guwahati or used after due permission from the concerned authority at the institutions having such facilities. The statistical analyses namely univariate and multivariable logistic regression analysis were carried out using SPSS (16.0 version). Validation of the models was also done through STATA 11.0. All these analyses were performed on HP 630 Laptop.