Chapter–5

Research Methodology

The Locale of Research

The local of the study was selected based on the following criteria.

Stage I:

Champhai District was selected purposefully because of its advantageous, geographical location and prespondence of school children of the age category we were looking for.

Stage II:

Champhai District is divided into 4 rural development blocks namely: Khawbung, Khawzawl, Ngopa and Champhai. For data collection, one block i.e. Champhai block is selected because of its location accessibility and four localities from it were selected subsequently.

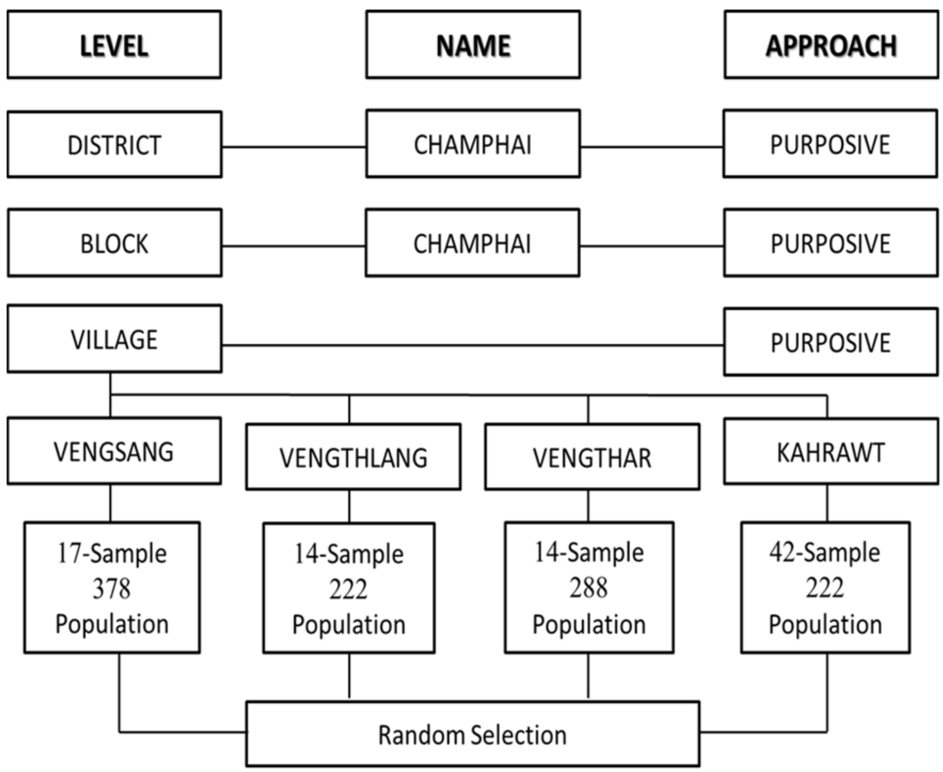
Stage III:

Four localities namely, Vengsang, Kahrawt, Vengthlang and Vengthar were selected to make the sample respondents representative geographically, culturally and socially.

Selection of the respondents:

Purposive and multi stage, purposive and random sampling method was followed here to select the respondent of 80 members. The respondent belonged to 6 – 14 age category attending primary school and middle school system. So far as the gender is concerned 50% of them are boys and 50% of them are girls. From Vengsang locality 17 children have been selected out of 378 children, from Kahrawt locality 42 children have been selected out of 222 children, from Vengthlang locality 14 children have been selected out of 288 children and lastly from Vengthar locality 7 children have been selected out of 201 children. This way the total size of respondent stood 80. This was rather a disproportionate random selection of respondent.

FLOW CHART ON SAMPLING METHOD:



ISBN: 978-93-85822-13-1

Time of Data collection:

The data was collected in the month of May, 2009.

Table 5.1: Variables and Empirical Measurement

	Variables	Empirical Measurement
X1	Age	Chronological age in terms of years.
X2	Education	Number of year of schooling.
X3	Parents Education	Summation of year of schoolings experienced both by mother and father.
X4	Family Size	Total number of members of the family under a single headship.
X5	Size of Homestead Land	Area and surroundings of a house structure expressed in decimal terms.
X6	Size of Agricultural Land	Total land under a single land possession and measured in terms of decimal.
X7	Agriculture Income	Total income derived from agriculture and expressed in terms of money. Income of agriculture is divided by the number of family size.
X8	Subsidiary Income	Income generating from subsidiary rural avocation divided by the number of family size. It is expressed in terms of money.
X9	Total Income	Summation of income from agriculture and other sources.
X10	Total Crop Yield	Total crop yields at different piece of land and expressed in kilogram.
X11	Home Consumption	The portion of crop yield consumed by the family members excluding the marketable surplus.
X12	Training	Variables measured by counting the number of training or field exposure the parents of the children have undergone.

ISBN: 978-93-85822-13-1

OPERATIONAL DEFINITION OF THE INDEPENDENT VARIABLES:

Age (X_1):

It denotes the chronological age, years and months elapsed since birth of the respondent. It was measured through counting the chronological age.

Education (X_2):

Education is the factor that has been conceived in terms of acquisition of knowledge and skill formality in school. The values ascribed the class in which he or she is studying.

Parents Education (X_3):

The parents' education has been measured in terms of year of schooling they underwent. The digital value of year of schooling was calculated for both the father and mother and the total value was divided by two to get the mean value of parent's education.

Family size (X_4):

It denotes the total no. of persons living in the respondents' house under a single household.

Size of homestead land (X_5):

It is the physical amount of the land mass under the ownership of the family land for the homestead. It is measured in terms of decimal.

Size of Agricultural land (X_6):

It is physical amount of land mass under the ownership of the family head for the production of crops. It is measured in terms of decimal.

ISBN: 978-93-85822-13-1

Agriculture income (X_7):

It is the income earned yearly by the family from agriculture products. The total income from agriculture was calculated by dividing the yearly income by the number of family members. It is measured in terms of money.

Subsidiary Income (X_8):

It is the income other than from agriculture. This is also calculated by dividing the monthly income by the number of family members. It is measured in terms of money.

Total income (X_9):

This is the total income of the family which was calculated by adding the income from agriculture and subsidiary. It is measured in terms of money.

Total yield (X_{10}):

It is the total yield of different crops of the respondents' family which was calculated yearly in kilogram.

Home consumption (X_{11}):

It is the quantity of vegetables which the family consumes from their own field and production.

Training (X_{12}):

It is the training which respondents' parents attend mostly in schools. It is calculated simply by the no. of times parents attend the meeting in school per year.

ISBN: 978-93-85822-13-1

Table 5.2: Dependent Variables:

Y1	:	Food Intake Volume
Y2	:	Calorie Consumption from Primary Food
Y3	:	Intake of High Value Food
Y4	:	Calorie Consumption from High Value Food
Y5	:	Total Calorie Consumption
Y6	:	Level of Sanitation
Y7	:	Nutritional Status

DESCRIPTION OF DEPENDENT VARIABLES

Food Intake Volume:

This is the amount of food a respondent takes per day. This includes primary food which is taken during lunch and dinner including morning and evening tea. It is calculated in gram.

Calorie Consumption from Primary Food:

This is the calorie consumption level of the respondent which is calculated from the primary food intake of respondents.

High Value Food:

This is the high value food intake of the respondents which include milk, meat and fruits. This is calculated by the number of times a respondent takes high value food per week divided by 7.

Calorie Consumption from High Value Food:

This is the calorie consumption level of the respondent from the high value food.

ISBN: 978-93-85822-13-1

Total Calories:

This is the total calorie consumed by the respondents. It is calculated by adding the calorie value from the food intake and high value food.

Level of Sanitation:

This is the sanitation level of the respondents family which includes toilet, bathroom and clothes. It has been divided into 3 levels such as;

Good – 3

Medium – 2

Bad – 1

After grading each of the respondents' toilet, bathroom and clothes according to the given levels, the three levels were added and again it is divided by 3 to get the result.

Statistical Tools Used

Mean:

The mean is the arithmetic average and is the result obtained when the sum of the value of the individuals in the data is divided by the number of individuals in the data (Pause and Sukhatme, 1967). Mean is the simplest and relatively stable measure of central tendency. The mean reflect and is affected by every score in the distribution. Thus extreme scores affect the mean.

For social action purpose, a mean may not provide a realistic picture of the situation. For example, the high income of a few big farmers may level off the poor income of the large number of marginal farmers.

ISBN: 978-93-85822-13-1

When the data are expressed in a frequency distribution (grouped), the mean is calculated by the formula:

$$\bar{X}=\frac{\sum f.x}{N}$$

Where, X=mean of the distribution

f= frequency of the class

x= class value or midpoint of the class interval

N= number of observations

Standard Deviation:

Standard Deviation is the square root of the arithmetic mean of the square of all deviation, the deviations being measured from the arithmetic mean of the distribution. It is commo9nly denoted by the symbol σ (sigma). It is less affected by sampling errors and is a more stable measure of dispersion. The standard deviation of the data grouped in the form of a frequency distribution is computed by the formula:

$$\sigma=\sqrt{\frac{\sum f.d2}{N}}$$

Where,

f= frequency of the class

d= deviation of the mid-value of the class from the population mean

N= total number of observations.

Co-efficient of Variation:

A measure of variation which is independent of the unit of measurement is provided by the co-efficient of variation. Being unit free, this is useful for comparison of variability between different populations. The co-efficient of

ISBN: 978-93-85822-13-1

variation is standard deviation expressed as percentage of the mean and is measured by the formula:

$$\text{Co-efficient of variance (C.V.)} = \frac{Standard\ Deviation}{Mean} \times 100$$

Correlation:

When an increase or decrease in one variant is accompanied by an increase or decrease in the other variant, the two are said to be correlated and the phenomenon is known as correlation. Correlation co-efficient(r) is a measure of relationship between two variables which are at the interval or ratio level of measurement and are linearly related. A person product moment (r) is computed by the formula:

$$rxy = \frac{N\sum \times Y - (\sum \times)(\sum Y)}{\sqrt{[N\sum \times 2 - (\sum \times)2][N\sum Y2 - (\sum Y2)]}}$$

Where, X and Y =original scores in variables X and Y

N= number of paired scores

$\sum \times Y$= each Y multiplied by its corresponding Y, then summed

$\sum \times$= sum of X score

$\sum \times^2$=each X squared, then summed

$(\sum \times)^2$= sum of X scores, squared

$\sum Y^2$= each Y squared, then summed

$(\sum Y)^2$= sum of Y scores, squared

$\sum Y$= sum of y score

The range of correlation co-efficient is between -1 to +1. This means that -1 is perfect negative correlation and +1 is perfect positive correlation. A perfect correlation is, however, achieved. An idea of negative and positive

ISBN: 978-93-85822-13-1

correlation is given here. If the number of errors increases with increase in typing speed, it indicates positive correlation. If the number of correct words decreases with increase typing speed it is indicating of negative correlation. A correlation co-efficient to be acceptable should be statistically significant. Otherwise no significant relationship exists between the variable.

Path Analysis:

The term was first introduced by the biologist Sewal Wright in 1934 in connection with decomposing the total correlation between any two variables in a causal system. The technique is based on a service of multiple regression analysis with the added assumption of the causal relationship between independent and dependent variable.

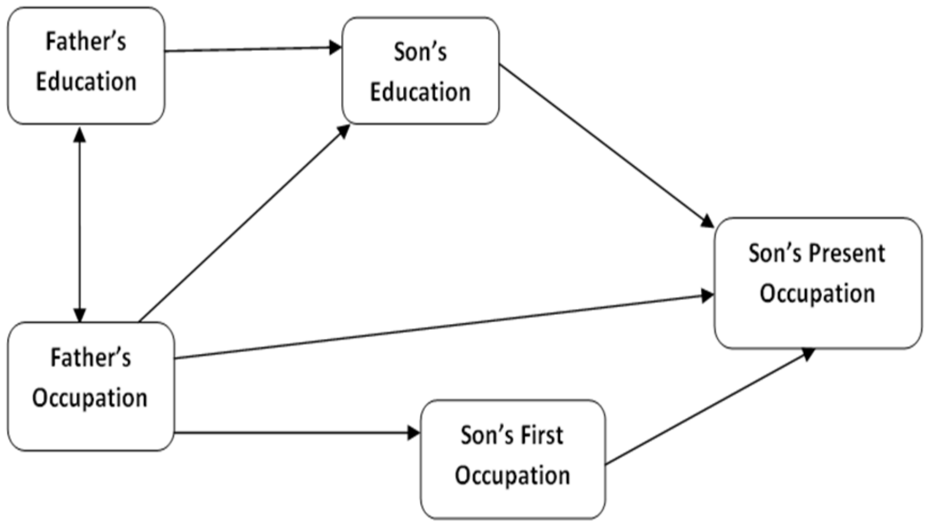
Path analysis makes use of standardized partial regression co-efficient (known as beta weights) was effect co-efficient. In linear additive affects are assumed, then through path analysis simple set of equations can be built up showing how each variable depend on preceding variable. The main principle of path analysis is that a correlation coefficient between two variables, or a gross or overall measure of empirical relationship can be decomposed in a series of parts: separate parts of influence leading through chronologically intermediate variable to which both the correlated variable have links.

The merit of path analysis in comparison to correlation analysis is that it makes possible the assessment of the relative influence of each antecedent or explanatory variables on the consequent or criterion variables by first

ISBN: 978-93-85822-13-1

making explicit the assumption, underlying the causal connections and then by elucidation the direct effect the explanatory variables.

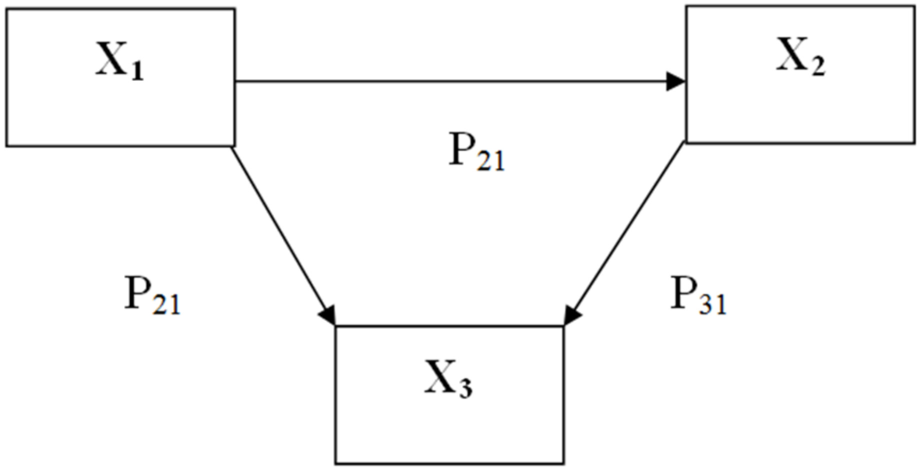
An illustrative path diagram showing inter relationship between father's education, father's occupation, son's first and son's present occupation can be shown as;



The use of the path analysis technique requires the assumption that there are linear additives, a symmetry relationship among a set of variables which can be measured at least on a quest interval scale. Each dependent variable is regarded as determined by the variable preceding it in the path diagram, and a residual variable defined as uncorrelated with other variables, is postulated to account for the unexplained portion of the variance in the dependent variable. The determining variables are summed for the analysis to be given (exogenous in the model).

ISBN: 978-93-85822-13-1

We may illustrate the path analysis technique in connection with a simple problem of testing a causal model with three explicit variables as shown in the following path diagram:



Path diagram (with three variables)

The structural equation for the above can be written as:

$X_1 = e_1$

$X_2 = P_{21}X_1 + e_2$ $= px+e$

$X_3 = P_{53}X_2 + P_{32}X_2 + e_3$

X1 and X variable are measured as deviation from their respective means. P_{21} may be estimated from the simple regression of X_2 on X_1, i.e.,$b_{21}X_1$ and P_{31}may be estimated from the regression of X on X_2 and X_1 as under:

$X= P_{31}X_1+b_{21}x_2$

Where, $b_{21}x_2$ means the standardized partial regression coefficient for predicting variable 1 when the effect of variable 2 is held constant.

ISBN: 978-93-85822-13-1

In path analysis the beta co-efficient indicates the direct of X_1 (j=1,2,3,...p) on the dependent variable. Squaring the direct effect yields the proportion of variance on the dependent variable Y which is due each of the number of independent variable X_1 (j=1,2,3...p). After calculating the direct effect one may obtain a summary measure of the total indirect of X_1 on the dependent variable Y by subtracting from the correlation coefficient r_{yxj} the beta co-efficient b i.e.

Indirect effect X_1 on $y=C_{jy}=r_{yxj}-b_1$

For all j=1, 2, 3...p

Discriminant Analysis:

Through discriminant analysis technique we can classify individuals or objects into one of two or more mutually exclusive and exhaustive groups on the basis of a set of independent variables. Discriminant analysis requires interval independent variables and a nominal dependent variable. For example, suppose that brand preference (say brand X and Y) is the dependent variable of interest and its relationship to an individual's income, age, education, etc. Is being investigated, and then we should use the technique of discriminant analysis. Regression analysis in such a situation is not suitable because the dependent variable is not intervally scaled. Thus discriminant analysis is considered an appropriate technique when the single dependent variable happens to be non-metric and is to be classified into two or more groups depending upon its relationship with several independent variables which all happen to be metric. The objective in discriminant analysis happens to be to predict an object's likelihood of

ISBN: 978-93-85822-13-1

belonging to a particular group based on several independent variables. In case we classify the dependent variable in more than two groups, then we use the name multiple discriminant analysis; but in case only two groups are to be formed, we simply use the term discriminant analysis.

In case only two groups of the individuals are to be formed on the basis of several independent variables, we can have a model like this

$$z_1=b_0+b_1X_{1i}+b_2X_{2i}+...+b_nX_{ni}$$

Where,

X_{ji}=the *i*th individual's value of the jth independent variable

b_j=the discriminant coefficient of the jth variable

z_i=the *i*th individual's discriminant score

z_{crit}=the critical value for the discriminant score

The classification procedure in such a case would be

If $z_i > z_{crit}$, classify individual *i* as belonging to Group-I

If $z_i < z_{crit}$, classify individual *i* as belonging to Group-II

When n (the number of independent variable) is equal to 2, we have a straight line classification boundary. Every individual on one side of the line is classified as Group-I and classification boundary is a 2-dimensional plane in 3 space and in general the classification boundary is an n-1 dimensional hyper-plane or n-space.

For judging the statistical significance between two groups, we work out the Mahalonobis statistic, D2, which happens to be a generalized distance between two groups, where each group is characterised by the same set of n variables and where it is assumed that variance-covariance structure is identical for both groups. It is worked out thus:

ISBN: 978-93-85822-13-1

$$D^2 = (U_1 - U_2)\nu^{-1}(U_1 - U_2)'$$

Where U_1=the mean vector for Group I

U_2=the mean vector for Group II

ν=the common variance matrix

By transformation procedure, this D^2 statistic becomes an F statistic which can be used to see if the two groups are statistically different from each other.

ISBN: 978-93-85822-13-1