Chapter 5

Research Methodology

In this chapter discussion on the methodology has been made to understand the concepts, methods and techniques, which are utilized to design the study, collect the information, analyze the data and interpret the findings for revelation of truth and formulation of theories. The entire discussion for easy understanding has been made under the following sub-heads.

5.1. Locale of research

5.2. Sampling design

5.3. Pilot Study

5.4. Variables and their measurements

5.5. Methods of data collection

5.6. Statistical tools used for analysis of data.

5.1. Locale of Research

The present study was conducted in two adjoining districts, Hooghly and Nadia. The village, Ghoshalia of Balagarh block in Hooghly district and the village, Maheswarpur of Chakdah block in Nadia district of the state West Bengal were selected for the study.

- The characters and the factors under study have been well discernible to this area;
- The researcher's close familiarity with respect to area, people, officials and local dialects;

- The ample opportunity to generate relevant data due to the close proximity of the area with the research and extension wing of the state Agriculture;
- The highly cooperative, responsive respondents;
- The profuse scope to get relevant information regarding energy consumption pattern, status of home condition, expenditure allotment, innovation proneness, food habits etc regarding agricultural technology;
- Experienced, well versed, venturesome and risk bearing farm entrepreneurs;
- Easy accessibility of the area;
- The study would help the researcher to conduct diversified extension programs and activities in future.

5.2. Sampling Design

The purposive as well as simple random sampling techniques were adopted for the present study. It may be termed as multistage random sampling procedure. The two districts ,Hooghly and Nadia, were considered purposively. Two blocks one in district Hooghly, Balagarh and another one Rautari from Nadia district were selected purposively for the study. Two villages one in block Chakdah, named Ghoshalia and another in Rautari block, named Maheswarpur were selected purposively for the study. 100 respondents, 50 from each village had been selected randomly for finale data collection.

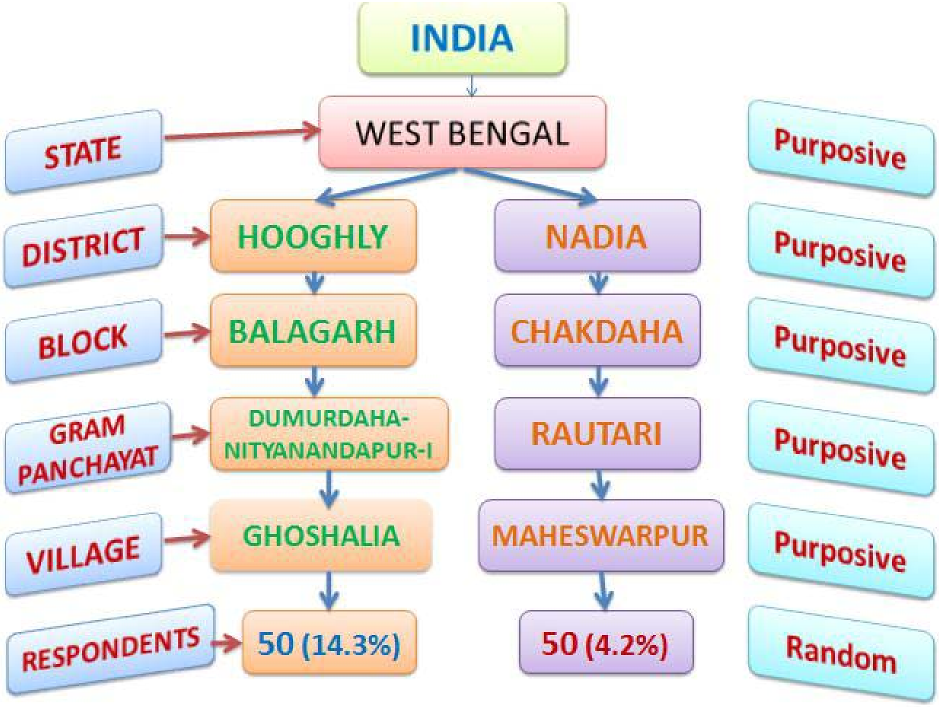

Figure 5.1 Flow chart of research locale

5.3. Pilot Study

A pilot study was conducted in the selected villages before constructing the data collecting devices. In course of this survey, informal discussion was carried out with some farmers, local leaders and extension agents of the localities. An outline of the socio-economic background of the farmers of the concerned villages, their opinion towards different types of technology socialization process, innovation-decision process, discontinuance, disagreement, conflict, rejection, dissonance, reinvention and confusion helped in the construction of reformative working tools.

The components of pilot study were:

- General information;
- Specific information;

- Prevalence of variables;
- Body languages of the prospective respondents;
- Access to physical location;
- The type, level and intensity of responsiveness;
- Related information including Agriculture.

5.4 Variables and Their Measurements

Several researchers pointed out that the behaviour of an individual was understood more in depth if one has the knowledge of some variables, which comprised the constructed world of reality within which an individual received the stimuli and acts. The socio personal, agro economic, socio-psychological and communication variables are such type of variables, which determine the behaviour of an individual. Appropriate operationalization and measurement of the variables help the researcher to land upon the accurate conclusion. Therefore, the selected variables for this study had been operationalised and measured in following manner. Variables in the present study have been categorized into two main categories.

- Independent variables/ Predictor variables;
- Dependent variables.

5.4.1 The Independent Variables

Age (X_1): In all societies, age is one of the most important determinants of social status and social role of the individual. In the present study, the number of years rounded in the nearest whole number of the respondents lived since birth at the time of interview, was taken as a measure of age of the farmer.

Education (X_2): - Education may be operationalized in the amount of formal schooling attained/literacy acquired by the respondent at the time of interview. Education is instrumental in building personality structure and helps in charging one‟s behavior in social life.

Gender ratio (X_3): - Gender was operationalized in the present study as number of male in family divided by the number of female in the same family.

Family size (X_4): - Family size was operationalized as the members in the individual family. In the present study only those members of the family considered, who were taking the meal in one chullah.

Family Education Status (X_5): -In the present study family education status is the average education achieved by the members of the family.

Innovation Index (X_6): -In the present study innovation index is taken as the total price of innovative technologies like mobile phone, motorbike, coloured TV, power tiller, hand tractor, tractor etc by the family members also by the number of innovative things.

Occupation (X_7): -Occupation of a person refers to regular activity performed for payment that occupies one‟s time. In the present study a scale was developed on 1-6 point scale. This scale consists of as follows.

Table 5.1 Occupation Measurement in 6 Point Scale

Sl. No.	Items	Scale (1-6)
1	Labour	
2	Artisan	
3	Business	
4	Independent	
5	Farming	
6	Services	

Family MIS (X_8): - Family MIS is the sum total of information received, information sent and information showed on TV or Radio or Newspaper or Magazine or Journal or through KVK scientist or other farmers divided by 3 (3 are the ways of information managed).

Cropping Intensity (X_9): - Cropping intensity has been operaionalized as the proportion of total annual cropped area to the size of holding expressed in percentage. The cropping intensity was calculated by the formula

× 100 %

Farm size (X_{10}): - Farm size is a measure of farm business. Operationally farm size may be defined as a tract of land possessed by an individual for the purpose of growing crops. Different research workers had tried to measure farms size in different ways. In the present study, actual area under cultivation in decimal divided by size of the family was taken as measure of farm size.

Expenditure Allotment (X_{11}): - In the present study per cent of expenditure incurred on farming of the total expenditure incurred annually by farm family. Total expenditure was calculated as follows

- Expenditure incurred on food annually.
- Expenditure incurred on Clothes annually.
- Expenditure incurred on Education annually.
- Expenditure incurred on Farming annually.
- Expenditure incurred on health annually.

100

Credit load (X_{12}): - Credit load of the farmers indicates that how much credit farmers have in outstanding. In the present study credit load of the farmers has been calculated as

Annual Income (X_{13}): - Annual income is the economic measurement of farmers‟ status. It was operationally defined as the gross income from all the viable sources of income in a single year. It was measured in terms of rounded of rupees. The gross income was constituted by the total income generated from agriculture, dairy, poultry, fishery enterprises, business and services. In the present study it has been calculated with the formula as follows.

Irrigation Index (X_{14}): - In the present study irrigation index was calculated in per cent as follows

× 100

Crop Diversity Index (X_{15}): - It is the measurement of farmers‟ diversified mind set up and his capability to handle more number of crops in one small farm holding. In this present study it is measured with the formula as follows.

Crop Energy Productivity (X_{16}): - It is the measurement of total amount of outcome from a cultivable land as crop yield, crop stubble, crop residue (straw or sticks) in terms of mega joule from the following websites-

www.fao.org/docrep/006/y5022e/y5022e04.htm

http://nutritiondata.self.com/fact

http://calorielab.com/foods/wheat/21

https://www.icac.org/projects/CommonFund/20_ucbvp/papers/01_dubey.pdf

Rice Straw

- 1 ton of Rice paddy produces 290 kg Rice Straw
- 290 kg Rice Straw can produce 100 kWh of power
- Calorific value = 2400 kcal/kg

Rice Husk

- 1 ton of Rice paddy produces 220 kg Rice Husk
- 1 ton Rice Husk is equivalent to 410- 570 kWh electricity
- Calorific value = 3000 kcal/kg
- Moisture content = 5 – 12%

Adoption Index (X_{17}): - It is the ratio between the numbers of adopted technology by the recommended technology (new variety of seed, fertilizers, pesticide, methods of cultivation etc.)

Size of Water Holding (X_{18}): - In this present study size of water holding is measured through following formula

$$\frac{\times}{\quad}$$

Cattle Holding Economics (X_{19}): - In this study Cattle holding economics is measured through the total amount of cost of maintenance per cattle rearing.

Cattle Energy Balance (Y_1): - It is defined as the difference between the energy equivalents of feed taken by the cattle and the energy equivalents of the output from cattle in the form of dung and milk per day per cattle.

Energy equivalence of Cow dung (Y_2): - It is defined as the total amount of cow dung used as bio fuel in the family, organic fertilizer in the field in terms of mega joule energy value.

Crop Energy Metabolism (Y_3): - It is defined as the difference between the energy equivalents of energy consumed by crop in the form of fertilizer, irrigation, ploughing, plant protection chemicals and the energy equivalents of the energy produced from the crop in the form of biological productivity (Crop Energy Productivity $X_{16)}$ per crop.

Energy Consumption in Farm Family (Y_4): -It is defined as the energy equivalents of the energy consumed by a household in different activities in the forms of electricity, dung cakes, fuel wood, kerosene, diesel, petrol, LPG per day per member.

Perceived Impact on Energy Consumption (Y_5): -It is defined as the impact of the energy consumption in the form of electricity, kerosene, petrol, diesel, dung cakes etc on the economy, society, occupation and ecology .It is calculated on 10 point scale.

Farmers' Energy Metabolism (Y_6): -It is defined as the subtraction between the total amount of energy intake i.e. food intake per day per person in terms of kilo joule and the total amount of energy output i.e. energy emission during farm operation, schooling, domestic working, travelling etc by a person in terms of kilo joule.

5.4.3 Method of calculating energy from raw data

- **Energy units and dimensions**

So far, we have discussed energy in qualitative terms. In order to proceed, we must discuss energy quantitatively. That means, we need units for measuring quantities of energy and related concepts. We use the International system of units (SI units), which is based on the dimensions and basic units in Table 5.2

Table 5.2 Basic SI units

Dimension	**Basic unit**	**Symbol**
Length	Meter	m
Mass	Kilogram	kg
Time	Second	s
electric current	Ampere	A
Temperature	Kelvin	°K

The unit of energy in this unit system is joule (J), and the unit of power is watt (W). These and many other units can be derived from the basic SI units. The relationship between some derived SI units and the basic SI units are represented in Table 5.3

Table 5.3 Derived SI units

Dimension	**Unit**	**symbol**
Area	square meter	m^2
Volume	cubic meter	m^3
Speed	meter per second	m/s
Acceleration	meter per second	m/s^2
Pressure	Pascal	Pa (=N/m)
volume flow	cubic meter per second	$m\ ^3/s$
mass flow	kilogram per second	kg/s
Density	kilogram per cubic meter	kg/m^3
Force	newton (*)	$N(=kg.m/s^2)$
Energy	joule (**)	J(=N.m)
Power	Watt	W (=J/s)
energy flux	watt per square meter	W/m^2
calorific value	joule per kilogram	J/kg
specific heat	joule per kilogram Kelvin	J/kg.K
Voltage	Volt	V (=W/A)

(*) The force exerted by a mass of 1 kg equals ca. 10 N.

(**) The energy required to lift 1 kg by 1 meter. Note that = W.s.

In some countries, or in a particular context, other units than SI units are also used. They can be converted into SI units, which are more convenient for calculations. The conversion of some non-SI units into SI units is given in Table 24, for energy.

Table 5.4 Conversion of non-SI units to SI units

Non-SI unit for energy	**Symbol**	**equivalence in SI-units**
Erg	Erg	10-7 J
foot pound force	ft.lbf	1.356 J
Calorie	Cal	4.187 J
Kilogram force meter	Kg f. m	9.8 J
British thermal unit	Btu	1.055 x 103 J
horsepower hour (metric)	hp.hr	2.646 x 106 J
horsepower hour (GB)	hp.hr	2.686 x 106 J
kilowatt hour	kWh	3.60 x 106 J
barrel oil equivalent	b.o.e.	6.119 x 109 J
ton wood equivalent	-	9.83 x 109 J
ton coal equivalent	Tee	29.31 x 109 J
ton oil equivalent	Toe	41.87 x 109 J
quad (PBtu)	-	1.055 x 1018 J
tera watt year	TWy	31.5 x 1018 J

5.4.4 Magnitudes of energy forms

Now we have introduced units for measuring energy, we can make quantitative comparisons and calculations. The following results give us some feeling of magnitudes of energy, as represented in different energy forms.

The examples are all equivalent to about 100 kJ;

- Radiation from the sun on the roof of a house (of ca. 40 m^2) in 2.5 s

- Energy released in burning 3.5 g coal or 2.9 g petrol; or the energy stored in 1/4 slice of bread
- A large object (1,000 kg) at a height of 10 m
- Energy produced by a windmill of 3 m diameter in a wind speed of 5 m/s (a breeze) during 20 minutes; or the energy stored in the mass of a car (1,000 kg) moving at 50 km/h heat emanated in cooling three cups of coffee (0.4 kg) from 80°C to 20° C; or the energy needed to melt 0.3 kg ice
- An iron flywheel of 0.6 m diameter and 70 mm thick, rotating at 1,500 revolutions per second
- Energy consumed by a 100 W electric light bulb in 17 minutes

As has been stated in before, energy conversions always imply energy losses. This leads us to the concept of efficiency, as follows. A quantity of energy in a certain form is put into a machine or device, for conversion into another form of energy. The output energy in the desired form is only a part of the Input energy. The balance is the energy loss (usually in the form of diffused heat). It means the converter has less than 100% efficiency.

The efficiency of an energy converter is now defined as the quantity of energy in the desired form (the output energy) divided by the quantity of energy put in for conversion (the input energy). The efficiency is usually expressed by the Greek letter. Hence:

$$\eta = \frac{\text{output energy}}{\text{input energy}}$$

Table 5.5 Energy equivalent values of some fuels

Fuel	Unit	tonnes of coal equivalent	tonnes of oil equivalent	barrels of oil equivalent	GJ (*)
Coal	Tone	1.00	0.70	5.05	29.3

firewood (**) (airdried)	Tone	0.46	0.32	2.34	13.6
kerosine (jet fuel)	Tone	1.47	1.03	7.43	43.1
natural gas	1000 m3	1.19	0.83	6.00	34.8
Gasoline	Barrel	0.18	0.12	0.90	5.2
gasoil/diesel	barrel	0.20	0.14	1.00	5.7

(*) Note that GJ/tonne is the same as MJ/kg.

(**) Note that the energy equivalent of wood can vary a factor 3 depending on the moisture content of the wood.

However, what we can achieve with an amount of energy depends very much on how the energy is utilized, that is, on the efficiencies of the energy converters applied.

Efficiencies can vary enormously for different converters, as we have seen in Section 9. The energy equivalent is then of limited use to us. In practice, when comparing sources of energy, we are more interested in the replacement value of the energy form. The latter Indicates how much of that energy form is required to do the same job (i.e. serve the same use) as another energy form or fuel. Again, as a reference, coal is sometimes used. The replacement value of an energy form is, then again, expressed in tee. However, this value will be different from the equivalent value of that energy form.

An easy way of comparing replacement values of different energy forms is by indicating how many units of the energy form (or fuel) can replace one kg of coal. We call this the replacement ratio of the fuel. Replacement ratios of some household energy forms compared with coal are given in Table 8, as taken from a particular survey. (Alternatively, a similar table could be made with oil as a reference.) It should be noted that the figures serve as an example only, as they depend on the actual efficiencies of the conversion techniques applied.

Table 5.6 Coal replacement ratio of some forms of energy

energy form or fuel	Unit	coal replacement ratio (kg coal per unit)
dung cake	Kg	0.30
vegetable waste	Kg	0.60
Firewood	Kg	0.70 - 0.95
soft coke	Kg	1.50
Charcoal	Kg	1.80
kerosene (lamp)	1	2.10
kerosene (stove)	1	5.20 - 7.00
Electricity	kWh	0.70

(The coal replacement ratio is the number of kg of coal which is required to effectively replace 1 unit of the energy form or fuel, under certain assumptions.)

Good examples of coal replacement are a kerosene lamp and a kerosene stove. The coal equivalent of kerosene was 1.47, which means that the heating value of 1 kg kerosene equals that of 1.47 kg coal. However, the coal replacement ratio for a kerosene lamp is 2.10, which means that 2.10 kg coal would be required to get as much light as from 1 kg kerosene. And the coal replacement ratio of a kerosene stove is around 6, which means that 6 kg coal is required to get as much heat in a pot as from 1 kg kerosene.

In Section 7, it was mentioned that the breakdown of energy flows is relevant for surveys and statistics. This is illustrated by the previous discussion of energy equivalence and energy replacement. We can add the primary energy resources of a particular region by adding the energy equivalences of all the various primary energy resources available. This will give us a rather theoretical figure, as it does not say what can be done with this amount of energy. We can also add, say, the consumption of final energy for a particular sector in a region, and work this out in a coal replacement value. Or we can consider, say, the amount of useful energy for particular end-uses, and express this in an oil (or coal) replacement value. For working out the replacement values, we should know the conversion methods and their efficiencies which are involved in the energy flow.

Note: The energy equivalents (calorific value) for different fuels, crop productivity has been taken from following websites.

www.energy.korea.com

www.mnre.gov.in

www.wikipedia.org/milk

www.ces.iisc.ernet.in

www.mospi.nic.in

5.5 Methods of Data Collection

Preparation of Interview Schedule

On the basis of the findings of pilot study a preliminary interview schedule was formed with the help of literature and by the assistance of Chairman of Advisory Committee. The interview schedule consisted of three major parts according to the specific objectives of the study.

Pre-testing of Interview Schedule

Pretesting or preliminary testing is the process of an advance testing of the study design after the schedule/questionnaire has been prepared. The object of pretesting is to detect the discrepancies that have emerged and to remove them after necessary modification in the schedule. It also helps to identify whether the questions are logically organized, the replies could properly recorded in the space provided for or there is any scope for further improvement. After conducting pretesting appropriate changes and modification of the interview schedule have been made. The individuals who responded in pretesting have been excluded in the final sample selected for the study.

Techniques of field data collection

The respondents were personally interviewed during puja vacation and summer vacation. The items were asked in Bengali as well as English version in a simple term so that the members could understand easily. The entries were done in the schedule by student investigator himself at the time of interview.

Construction of Schedule after Pre-Testing

The draft schedule for collection of data, incorporating the tools and techniques of different variables was presented twice each time on contact farmers. The quantification was done for each and every variable after operationalizing them. Before final data collection, entire schedule was pretested for elimination, addition and alteration with non-sample respondents of the study area. In pre-testing, care was taken not to include respondents who were selected as sample for final interview. On the basis of experience in pre-testing, appropriate changes in the construction of item and their sequence were made. The schedule was then finalized and multiplicitied. The final form of the schedule is given in the appendix.

5.5.1 Field Data Collection

The primary data in the present study were collected directly from the farmers with the help of structured schedule through personal interview methods. Only the functional head of the household were taken as respondents for the study. The personal interview method was followed during the month of November 2012 to January 2013 to collect the relevant information form targeted respondents. In each village, before starting the interview, a few days were devoted to establish rapport with the respondents. The schedule was administered to the respondents in local language and the responses were recorded in English on the schedule. The interview was carried out by the researcher himself.

5.6 Statistical Analysis and Interpretation of Data (Analytical Tools)

After collection of data, data were processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. Processing implies editing, coding, classification, and tabulation of collected data. The Statistical techniques and tools used in the present study.

- **Mean**

Measure of central tendency (or statistical averages) tells us the point about which items have a tendency to cluster. Such a measure is considered as the most representative figure for the entire mass of data. Measure of central tendency is also known as statistical average. Mean, median and mode are the most popular averages. Mean, also known as arithmetic average, is the most common measure of central tendency and may be defined as the value, which we get by dividing the total of the values of various given items in a series by the total number of items. We can work it out as follows

$$\text{Mean or } (x) = \frac{\sum}{\ } = \frac{{}_1+{}_2+{}_3+\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots}{\ }$$

Where,

(x) = The symbol we use for mean (pronounced as x bar)

∑ = Symbol for summation

x_i = Value of the i^{th} item X , I = 1, 2, ……………………….n

N = Total number of items.

Mean is the simplest measurement of central tendency and is a widely used measure. Its chief use consists in summarizing the essential features of a series and in enabling data to be compared. It is a relatively stable measure of central tendency. But it suffers from some limitations *viz.* it is unduly affected by extreme; it may not coincide with actual value of an item in a series, and it may lead to strong impressions, particularly when the item values are not given with the average. However, mean is better than other average, especially in economic and social studies where direct quantitative measurements are possible.

- **Standard Deviation**

Standard deviation is the most widely used measure of dispersion of a series and is commonly denoted by the symbol **σ** (pronounced as sigma) Standard deviation is the square root of the arithmetic mean of the square of the deviations, the deviations being measured from the arithmetic mean of distribution. It is less affected by sampling errors and is more stable measure of dispersion. It is worked out as follows,

$$\text{Standard deviation } (\boldsymbol{\sigma}) = \frac{\overline{\sum(\ -\ )^2}}{}$$

- **Coefficient of Variation**

A measure of variation which is independent of the unit of measurement is provided by Coefficient of variation. Being unit free, this is useful for computation of variability between different populations. The Coefficient of variation is standard deviation expressed as percentage of the mean and is measured by the formula.

$$V = \frac{(\ )}{} \times 100$$

- **Coefficient of Correlation**

When an increase or decrease in one variable is accompanied by an increase or decrease in other variable, the two are said to be correlated and the phenomenon is known as correlation. Correlation coefficient (r) is a measure of the relationship between two variables, which are at the interval or ratio level or measurement and are linearly related. A Karl Pea‟rson coefficient of correlation also known as product moment „r‟ is computed by the formula.

$$= \frac{\sum\ -(\sum\ )(\sum\ )}{\sqrt{[\sum\ ^2 - (\sum\ )2][\sum\ ^2 - (\sum\ )2]}}$$

Where,

x and y = Original scores in variables x and y

N = Number of paired scores

$\sum$ = Each x multiplied by its corresponding y, then summed

$\sum$ = Sum of x scores

$\sum 2$ = Each x squared, then summed

$(\sum)2$ = Sum of x scores, squared

$\sum$ = Sum of y scores

$\sum 2$ = each y squared, then summed

$(\sum) 2$ = Sum of y scores, squared

This coefficient assumes the following;

- That there is linear relationship between the two variables;
- That the two variables are causally related which means that one of the variable is independent and other is dependent and;
- A large number of independent causes are operating in both variables so as to produce a normal distribution.

The value of „r‟ lies between +1 to -1. Positive values of r indicate that positive correlation between the two variables (i.e. changes in both variables take place in the same direction), whereas negative values of ‟„r indicate negative correlation i.e. changes in the two variables taking place in opposite direction. A zero value of „r‟ indicates that there is no association between the two variables. When r (+) 1, it indicates perfect positive correlation and when it is (-) 1, it indicates perfect negative correlation, meaning thereby that variations in independent variable (x) explain 100 per cent of the variations in the dependent variable (y). We can also say that for a unit change in independent variable, if there happens to be constant change in the dependent variable in the same direction, the correlation will be termed as perfect positive. But if such change occurs in the opposite direction, the correlation will be termed as perfect negative. The value of „r‟‟ nearer to +1 or -1 indicates high degree of correlation between the two variables.

- **Regression**

The correlation coefficient only expresses association and by itself tells nothing about the causal relationships of the variables. Thus, purely from the knowledge that two variables x and y are correlated, we cannot say whether variation in x is the cause or the results from mutual dependence of the two variables or from common causes affecting both of them. Similarly, the mere existence of a high value of correlation coefficient is not necessarily of an underlying relationship between the two variables.

The underlying relation between y and x in a bivariate population can be expressed in the form of a mathematical equation known as regression equation and is said to represent the regression of the variable y on the variable x. (Panse and Sukhatme, 1967)

If y is the dependent variable and x is the independent variable, then the linear regression equation can be written as

$$y = a + bx$$

The values of a and b can be obtained by the method of least squares which consists of minimizing the expression

$\sum(y_i - a - bx_i)^2$ with respect to a and b.

The values of a and b are

a = y [redacted] bx –

$$b = \frac{\sum \; - \frac{\sum xi\ (\sum yi\ )}{}}{\sum {}^{2} - \frac{(\sum\ )^{2}}{}}$$

The regression equation can now be wri~~t~~ten a~~s~~

y = y – bx +bx

y - y =b(~~x~~ - x) –

Where b is the regression coefficient

- **Stepwise Multiple Regression**

Stepwise regression is a variation of multiple regressions which provides a means of choosing independent variables that yield the best prediction possible with the fewest independent variables. It permits the user to solve a sequence of one or more multiple linear regression problems by stepwise application of the least square method. At each step in the analysis, a variable is added or removed which results in the greatest production in the error sum of squares (Burroughs Corporation, 1975).

According to Drapper and Smith (1981), the method of stepwise multiple regression analysis is to insert variables in turn until the regression equation in satisfactory. The order of insertion is determined by suing the partial correlation coefficient as a measure of the importance of variables not yet in the equation.

The program, according to Burroughs Corporation (1975), first forms a correlation matrix, finds the best predictor (the independent variable having the highest correlation with criterion variable) and performs a regression analysis with this predictor. Then, the second best predictor (independent), and so on. At any given step, the group of predictors being used is not necessarily the best group of that size (i.e. the particular group of independent variables does not necessarily have the highest multiple correlation with the criterion that any group of this size does). Rather, this group contains the variables that have the highest individual correlation with the criterion.

Significance of variable that is being considered for entrance into the regression equation is measured by the F-statistic. If F is too small (less than F „include“), the variable is not added to the regression equati on. Include statement establishes the minimum value of the F-statistic required for the inclusion of a variable in the regression equation. In the example which follows, the F-value for inclusion was 0.01.

Significance of variables already in the regression equation may change as new variables are entered. This significance of the variables currently in the equation is also measured by the F-statistic. If F is too small (less than F „delete“), the variable is not added to the equation. Delete establishes the value of the F-statistic below which the variable is deleted from the regression equation. Here, the F-value for deletion was 0.005.

The „tolerance‟ level specified is used as control of degeneracy occurs when a variable entered into the equation is a linear combination of variables already present in the equation. Tolerance statement establishes the maximum value a pivoted element may attain while still allowing its associated variable to be brought into equation. A variable is not brought into the regression equation if its associated pivoted element is below the specified tolerance level, which was 0.001 in the present example.

- **Path Analysis**

The term „path analysis‟ was first introduced by the biologist, Sewall Wright in 1934 in connection with decomposing the total correlation between any two variables in a causal system. The technique of path analysis is based on a series of multiple regression analysis with the added assumption of causal relationship between independent and dependent variables. Path analysis makes use of standardized partial regression coefficient (known as beta weights) as effect coefficients. In linear additive effects are assumed, and then through path analysis a simple set of equations can be built up showing how each variable depends on preceding variables. The main principle of path analysis is that any correlation coefficient between two variables, or a gross or overall measure of empirical relationship can be decomposed into a series of paths: separate paths of influence leading through chronologically intermediate variable to which both the correlated variables have linked.

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 variable on the consequent or criterion variables by first making explicit the assumptions underlying the causal connections and then by elucidating the indirect effect of the explanatory variables.

- **Factor Analysis**

Factor analysis is a very useful and popular method of multivariate research technique, mostly used in social and behavioural sciences. This technique is applicable when there is a systematic interdependence among a set of observed or manifest variables, and the researcher is interested in finding out something more fundamental or latent which creates this communality (commonness). For example, we may have data on farmers‘‘ education, occupation, land, house, farm power, material possession, social participation etc. and want to infer from these some factor relating to social status, which shall summarize the communality of all the variables.

According to Kothari (1996), Factor analysis seeks to resolve a large set of measured variables in terms of relatively few categories, known as factors. This technique allows the researcher to group variables into factors (based on correlation between variables), and the factors so derived may be treated as new variables (often termed as latent variables) and their grouped into the factor. The meaning and name of such new variable is subjectively determined by the researcher.

Since the factors happen to be linear combinations of data, the coordinates of each observation or variables is measured to obtain what are called factor loadings. Such factors loadings represent the correlation between the particular variable and the factor, and are usually placed in a matrix of correlations between the variable and the factors.

- **Concepts Used In Factor Analysis**

Some important concepts used in factor analysis are explained, following Kothari (1996).

Factor A factor is an underlying dimension that accounts for several observed variables. Factor is a hypothetical construct or classification. There may be one or more factors, depending upon the nature of the study and the number of variables involved in it.

Factor loadings: Factor loadings are those values which explain how closely the variables are related to each one of the factors discovered. Factor loadings work as key to understanding what the factors mean. It is the absolute size (rather the signs, plus or minus) of the loadings that is important in the interpretation of a factor.

Communality (h^2) Communality, represented by h^2 , shows how much of each variable is accounted for by the underlying factor taken together. A high value of communality means that not much of the variable is left over after whatever the factors represent is taken into consideration.

Eigenvalue (or latent root): The sum of squared values of factor loadings relating to a factor is referred to as eigenvlue or latent root. Eigenvalue indicates the relative importance of each factor in accounting for the particular set of variables being analyzed.

Rotation: Rotation reveals different structures in the data and provides meaning to the results of factor analysis. There are different types of rotations such as orthogonal rotations, oblique rotations, varimax rotation etc. One has to select a rotation appropriate to the study. For the present study varimax rotation has been used.

- **Principal Component Analysis**

There are several methods of factor analysis. The method of Principal Component Analysis which is widely used is discussed here.

The principal component analysis extracts m-eigenvectors (principal component axes) and corresponding m-eigenvalues (the variance measured along the eigenvector), from m x m symmetrical matrix of correlation. The eigenvectors obtained from this principal component analysis are all orthogonal (i.e. inter-column correlations are near zero). The eigenvalues account for all of the original data variances in decreasing order such that each has variance or eigenvalue less than the previous ones. The total of the eigenvalues ($_1$ + $_2$ + ⋯ … … … . + ,) which is the same as the sum of the variances constituting the diagonal or trace of the correlation matrix before transformation. The principal components are then converted into factors by multiplying each element of the principal components or eigenvectors (v) by the square – root of the corresponding eigenvalues ($^{1\ 2}$.). Factors, thus, besides the direction also represent the variances.

The analysis calls for the selection of a minimum number of meaningful and useful factors, considerably fewer in number than the original variables, which will account for most of the variances in the data set and therefore, convey the same information. Various criteria for selection of suitable factors are available. Kaiser (1958) and others have recommended retaining all those eigenvalues, which have values more than one.

Next step is to remove the noise imposed by (m - p) unnecessary axes. To accomplish this, p-orthogonal reference axes or factors are routed about the origin to positions such that the variance of the loading from each variable onto each factor axis is either extreme (±1) or zero. This maximization of the range of the loadings was performed by using Kaiser‟s Varimax criterion. Scanning through each factor column for large absolute values in the varimax matrix will reveal a few variables with significantly high loadings and many others with insignificantly loadings. The column showing communality $\sum h_j^2$) is the total amount of variance of each variable retained in the factors, and is computed by summing the squares of the elements of the factors in each row of the varimax matrix. Fairly high communality of each variable implies the appropriateness of the model adopted, for the study. The last step involved meaningful interpretation of the factors.

- **Canonical Correlation Analysis**

A canonical correlation is the correlation of two canonical (latent) variables, one representing a set of independent variables the other a set of dependent variables. Each set may be considered a latent variable based on measured indicator variables in its set. The canonical correlation is optimized such that the linear correlation between the two latent variables is maximized. Whereas multiple regressions are used for many to one relationship, canonical correlation is used for many to many relationships. There may be more than one such linear correlation relating the two set of variables, with each such correlation representing a different dimension by which the independent set of variables is related to the dependent set. The purpose of canonical correlation is to explain the relation of the two sets of variables not to model the individual variables.

Analogous with ordinary correlation, canonical correlation squared is the per cent of variance in the dependent set of variables along a given dimension (there may be more than one). In addition to asking how strong the relationship is between two latent variables, canonical correlation is useful in determining how many dimensions are needed to account for that relationship. Canonical correlation finds the linear combination of variables that produces the largest correlation with second set of variables. This linear combination or “root” is extracted and the process is repeated for the residual data, with the constraint that the second linear combination of variables must not correlate with the first one. The process is repeated until a successive linear combination is no longer significant.

Canonical correlation is a member of multiple general linear hypothesis (MLGH) family and shares many of assumptions of multiple regression such as linearity of relationship, homoscedasticity (same level of relationship for the full range of the data), interval or near interval data, untruncated variables, proper specification of model, lack of high multicolinearity and multivariate normality for purpose of hypothesis testing.

Often in applied research, scientist encounter variables of large dimensions and are faced with problem of understanding dependency structures reduction of dimensionalities, construction of a subset of good predictors from the explanatory variables etc. Canonical correlation analysis provides us with a tool to attack these problems.

- **Some comments on the Canonical Correlation**

There could be a situation where some of variables have high structure correlation even though their canonical weights are near zero. This could happen because the weight is partial coefficient whereas the structure correlations (canonical factor loading) are not: if a given variable share variance with other independent variable entered in the linear combinations of variables entered in the linear combinations of variables used to create a canonical variable, its canonical coefficient (weight) is computed based on the residual variance it can explain after controlling for these variables. If an independent variable is totally redundant with another independent variable, its partial coefficient (canonical weight) will be zero. Nonetheless, such a variable might have high correlation with canonical variable (that is high structures correlation) have to do with the sample overall correlation of the original variable with the canonical variable.

Canonical correlation is not a measure of the per cent of variance explained in the original variables. The square of the structure correlation is the per cent of variance in a given original variable accounted for by a given canonical variable on a given (usually the first) canonical correlation. Note that the average per cent of variance explained in the original variable by a canonical variable (the mean of the squared structure correlation for the canonical variable) is not all the same as the canonical correlation, which has to do with the correlation between the weighted sums of the two sets of variables. Put another way the canonical correlation does not tell us how much of the variance in the original variable is explained by the canonical variable instead, that is determined on the basis of the squares of the structure correlation.

Canonical coefficient can be used to explain with which original variable a canonical correlation is predominantly associated. The canonical coefficient are standardized coefficient and (like beta weights in regression). The magnitude can be compared. Looking at the columns in SPSS output which list the canonical coefficient as columns and the variable is a set of variables as rows, some researches simply not variable with the highest coefficient to determine which variable are associated with which canonical correlation and use this as the basis for inducing the meaning of the dimension represented by canonical correlations.

- **Redundancy in Canonical Correlation Analysis**

Redundancy is the per cent of variance in one set of variable accounted for by the variate of the other set. The researcher wants high redundancy indicating that independent variate accounts for a high per cent of variance in the dependent set of original variables. Note that this is not the canonical correlation squared which the per cent of variance in the dependent variate is accounted for the independent variate.

- **Discriminant Function Analysis**

Discriminant function analysis undertakes the same task as multiple linear regressions by predicting an outcome. However, multiple linear regressions is limited to cases where the dependent variable on the Y axis is an interval variable so that the combination of predictors will, through the regression equation, produce estimated mean population numerical Y values for given values of weighted combination of X values. But many interesting variables are categorical such as political party voting intensions, migrant/non-migrant status, making a profit or not, holding a particular credit card, owning, renting or paying a mortgage for a house, employed/unemployed, satisfied versus dissatisfied employees, which customers are likely to buy a product or not buy, what distinguishes stiller Bean clients from Gloria Beans clients, whether a person is a credit risk or not, etc.

Discriminant analysis is used when:

a) The dependent is categorized with the predictor‟s IV at interval level such as age, income, attitudes, perceptions and years of education, although dummy variables can be used as predictors as in multiple regressions logistic regressions IV‟s can be any level of measurement.

b) There are more than two DV categories, unlike logistic regression, which is limited to dichotomous dependent variables.

- **Discriminant Analysis linear equation**

Discriminant analysis linear equation involves the determination of a linear equation like regression that will predict which group the case belongs to. The form of the equation or function is

= 1 1 + 2 2 + 3 3 = ⋯ … … … … … … … … … +

Where, D = Discrimination function;

V = The discriminant coefficient or weight for that variable.

X = Respondent‟s score for that variable.

a = a constant

i = the number of predictor variable

This function is similar to a regressions equation or function. The V‟s are unstandardized discriminant cefficient analogue to the „b‟s in the regressions equation. These V‟s maximize the distance between the mean of the criterion (dependent) variable. Standardized discriminant coefficient can also be used like beta weight in regressions. Good predictors tend to have large weights. What is needed, this function to do is maximize the distance between the categories; i.e. come up with an equation that has strong discriminatory power between groups. After using an existing set of data to calculate the discriminant function and classify cases, any new cases can then be classified. The number of discriminant function is one less the number of groups. There is only one function for the basic two group discriminant analysis.

A discriminant score is a weighted linear combination (sum) of the discriminating variables.

Assumptions of discriminant analysis:

a) The observations are random sample;

b) Each predictor variable is normally distributed;

c) Each of the allocations for the dependent categories in the initial classification are usually classified;

d) There must be at least two groups or categories in the initial classification are correctly classified;

e) There must be at least two groups or categories, with each case belonging to only one group so that the groups are mutually exclusive and collectively exhaustive (all cases can be placed in a group);

f) Each group or category must be well defined, clearly differentiated from any other (group) and natural. Putting a median split on an attitude scale is not a natural way to form group. Partitioning quantitative variables is only justified if there are easily identifiable gaps at the points of division.

g) For instance, three groups taking their available levels of amount of housing loan;

h) The groups or categories should be defined before collecting the data;

i) The attributes used to separate the group should discriminate quite clearly between the groups so that group or category overlap is clearly non-existent or minimal;

j) Group sizes of the dependent should not be grossly different and should be at least five times the number of independent variables.

There are several purposes of Discriminant analysis;

i. To investigate differences between groups on the basis of the attributes of the cases, indicating which attributes contribute most to group separation. The descriptive technique successively identifies the linear combinations of attributes known as canonical discriminant functions (equation) which contribute maximally to group separation.

ii. Predictive discriminant analysis addresses the question of how to assign new cases to groups. The discriminant analysis function uses a person"s scores on the predictor variables to predict the category to which the individual belongs.

iii. To determine the most parsimonious way to distinguish between groups.

iv. To classify cases into groups, statistical significance tests using chi-square enable you to see how well the function separates the groups.

v. To test theory whether cases are classified as predicted.

vi. Discriminate analysis creates an equation which will minimize the possibility of misclassifying cases into their respective groups or categories.

The aim of the statistical analysis in discriminant analysis is to combine (wieght) the variable scores in same way so that a single new composite variables, the discriminant score is produced. One way of thinking about this is in terms of a food recipe, where changing the proportions (weights) of the ingredients will change the characteristics of the finished cakes. Hopefully the weighted combinations of ingredients will produce two different types of cake.

Similarly, at the end of the discriminant process, it is hoped that each group will have a normal distribution of discriminant scores. The degree of overlap between the discriminant score distribution can then be used as a measure of the success of the technique, so that, like the different types of cake mix, we have two different types of groups

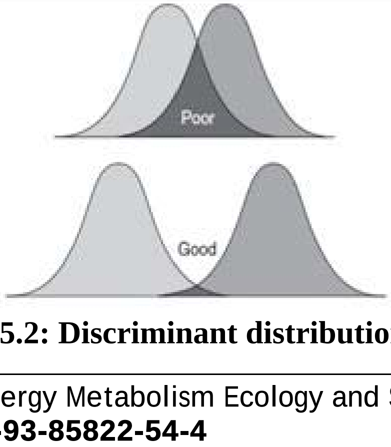

Fig. 5.2: Discriminant distribution

The top two distributions in figure overlap too much and do not discriminate too well compared to the bottom set. Misclassification will be minimized in the lower pair, whereas many will be misclassified in the top pair.

Standardizing the variables ensure that scale differences between the variables are eliminated. When all variables are standardized, absolute weights (i.e. ignore the sign) can be used to rank variables in terms of their discriminating power, the largest weight being associated with the most powerful discriminating variables with large weight are those which contribute mostly to differentially the groups.

As with most other multivariate methods, it is possible to present a pictorial explanation of the technique the following example uses a very simple data set, two groups and two variables. If scatter graphs are plotted for scores against the two variables, distributions like those in figure

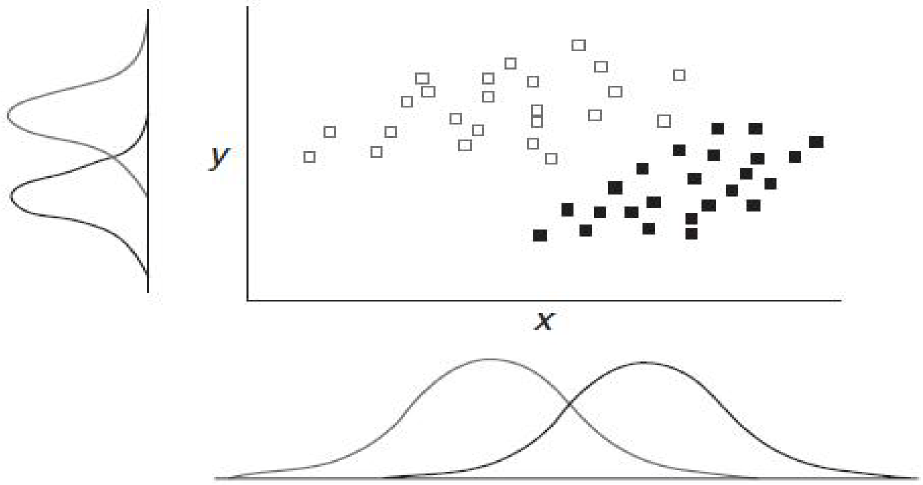

Fig. 5.3 Scatter graph displaying distribution

The new axis represents a new variable which is a linear combination of x and y i.e. it is a discriminant function (Fig. 4.3) obviously, with more than two groups or variables this graphical method becomes impossible.

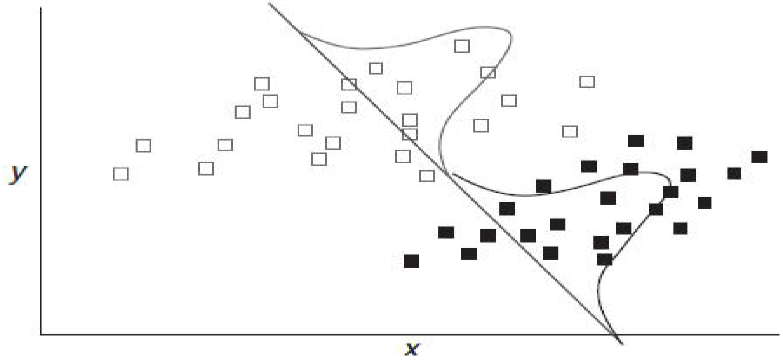

Fig. 5.4: New axis creating greater discrimination

Clearly, the two groups can be separated by their two variables, but there is a large amount of overlap on each single axis (although the y variable is the „better‟ discriminator). It is possible to construct a new axis which passes through the two group centroids (means‟) such that the groups do not overlap on the new axis.

In a two-group situation predicted membership is calculated by first producing a score of D for each case using the discriminant function. Then cases with D values larger are classified into the other group. SPSS will save the predicted group membership and D scores as new variables.

The group centroid is the mean value of the discrimination score for a given category of the dependent variable. There are as many centroids as there are groups or categories. The cut-off is the mean of the centroids. If the discriminant score of the function is less than or equal to the cut-off the case is classed as 0, whereas if it is above, it is classed as 1.

- **Stepwise Discriminant Analysis**

Discriminant analysis uses a collection of interval variables to predict a categorical variable that may be dichotomy or have more than two values. The technique involves finding a linear combination of independent variables (predictors) – the discriminant function- that creates the maximum difference between group memberships in the categorical dependent variable

In the present study for the stepwise discriminant analysis, canonical disrciminant function coefficients have been used. Stepwise discriminant analysis, like its parallel in multiple regressions, is an attempt to find the best set of predictors. It is often used in exploratory situation to identify those variables from among a large number that might be used later in a more rigorous theoretically driven study. In a stepwise discriminant analysis, the most correlated independent is entered first by the stepwise program, and then second until an additional dependent adds no significant amount to canonical R squared. The criteria of adding or removing are typically the setting of critical significance level for „F‟ to remove. These are unstandardized coefficient (b) and used to create the discriminant function (equation). It operates just like the regression equation.

The discriminant function coefficient „b‟ or standardized form „beta‟ both indicates the partial contribution of each variable to the discriminant function controlling for all other variables in the equation. They can be used to asses each ‟X‟sI (in the present study) unique contribution to the discriminant function and therefore provide information on the relative importance of each variable. If there are any dummy variables as in regression, individual „beta‟ weights cannot be used and dummy variables must be assessed as a group through hierarchical discriminant analysis running the analysis, first without the dummy variables then with them. The difference is squared canonical correlation indicates the explanatory effect on the set of dummy variable.

- **Canonical Discriminant Function**

Canonical Dicriminant Coefficient Table:

The understanderdized coefficients (b) are used to create the dicriminant function (equation). It operates just like regression equation. The discriminant function coefficient b or standardized farm beta both indicates the partial contribution of each variable to the discriminant functions controlling for the discriminant function and therefore provide information on the relative importance of each variable. If there are any dummy variables as in regression, individual beta weights cannot be used and dummy variables must be assessed as a group through hierarchical discriminant analysis running the analysis first without the dummy variables then with them. The difference is squared canonical correlation indicates the explanatory effect on the set of dummy variable.

Group Centroid Table

A further way of interpreting discriminant analysis results is to describe each group in terms of profile using the group means of the predictor variables. These group means are called centroid. These are displayed in group centroid tables.

Wilks' Lamda Table

This table reveals that all the predictors add some predictive power to the discriminant function as all are significant with p< .000.