Chapter–5

Research Methodology

Research methodology is the structural configuration of the study for conducting the research within the framework of the objectives. It includes methods, tools, techniques and approaches. For any research work, methodology furnishes the building block and backbone of the process of enquiry and reasoning, data generation and processing and so on.

Locale of Research

The study was conducted taking samples from the North-Eastern States of Arunachal Pradesh, Manipur, Meghalaya Mizoram and Nagaland. The states, districts, blocks and villages were selected purposively as tribal of different cultures are dominating in these areas and also because of the convenient and easier accessibility of the researcher. The states selected for this study are hilly regions with rich intercultural diversities which made the characters and the factor under study well discernible.

Pilot Study

Pilot study is a preliminary study conducted on a limited scale before the original studies are carried out in order to gain some primary information on the basis of which main project would be planned and formulated. Before

selection of the research topic, pilot study was conducted in Umiam, Meghalaya and Langol, Manipur. Extension personnel from KVK, Agriculture Officers and other friends from the selected states were contacted before the final data was collected to gain some information and the situation of the selected area.

Sampling Design

The states, districts, blocks and villages were selected purposively while the respondents were selected randomly. Out of the 17 districts of Arunachal Pradesh, three districts such as Lower Subansiri, Upper Subansiri and Papum Pare were selected. Ziro-I, Ziro-II and Raga were the three blocks selected and Hari, Bulla, Yachuli and Panjing were the villages selected. From the district of Upper Subansiri, blocks Daporijo and Dumporijo were selected. Paching colony, Lida and Digak were the villages selected. Itanagar and Naharlagun were the blocks and villages selected respectively from Papum Pare. 60 respondents were selected from the state of Arunachal Pradesh. From the state of Manipur three districts were selected out of the nine districts. From Churachandpur district, Tuiboung and Lanva blocks were selected and Molnom, Saipum and Hlahvom were the villages selected for the study. From Senapati District, Paomata and Tadubi were the blocks selected, Tungjoy, Kaibi and Maikhel were the villages selected and from Tamenglong district, Nungba block was selected, Kambiron was the village selected for the study. A total of 71 respondents were selected from the state of Manipur.

In Meghalaya, out of the 11 districts, three districts were selected namely; West Garo Hills, East Jaintia Hills and Rihbhoi district. From West Garo –

ISBN: 978-93-85822-06-3

Hill district, Rongram block and Gambigre block were selected and Chasingre and Damalgre were the villages selected from these blocks respectively. And from East Jaintia Hills, Khleihriat block and Bapung village were selected. From Rihbhoi District, Umsning block and Umiam village were selected. A total of 24 respondents were selected from Meghalaya. From the state of Mizoram, Champhai District was selected. Champhai block and Kahrawt village were selected for the stusy. A total of 10 respondents were selected from this area. From Nagaland, out of 11 districts, two districts were selected. From Kohima district, three blocks namely; Kohima, Tseminyu and Chiephobozou were selected. Chedema Model village and Chadema village were selected under Kohima block, Chunlikha and Gukhanyu villages were selected under Tseminyu block and Changsu New and Changsu Old were selected under Chiephobozou block. From Wokha district, two blocks, Wokha and Bhandari were selected. Yanmhon New and Yamnhon Old were the villages selected under Wokha bock. Sutemi and Zaphumi villages were selected under Bhandari district.

The interview schedule has been prepared with consultation of the interpreters of all states except in Mizoram so that relevant questions relating to the study can be selected and arranged logically in the interview schedule.

ISBN: 978-93-85822-06-3

Sampling Design at a Glance

State with total no.of districts	Districts Selected	Blocks Selected	Villages Selected	Name of tribes	Sample size	Selection procedure
	Lower Subansiri (Total Blocks=3)	Ziro-I Ziro II Raga	Hari Bulla Yachuli Panjing	Apatani Nishi Hills Miri	5 6 10 8	Districts, Blocks and Villages were selected purposively according to the majority of the tribes and according to the convenience of accessibility whereas respondent had been selected randomly with the assistance of interpreter
Arunachal Pradesh (17)	Upper Subansiri (Total Blocks=9)	Daporijo Dumporijo	Paching Colony Lida Digak	Tagin Galo	7 4 4	
	Papum Pare (Total Blocks=6)	Itanagar	Naharlagun	Adi Monpa	8 8	
Total	**3**	**6**	**8**	**7**	**60**	

ISBN: 978-93-85822-06-3

Manipur (9)	Churachandpur (Total Blocks=10)	Tuiboung Lanva	Molnom Saipum Hlahvom	Kuki Hmar Gangte Paihte Vaiphei	9 8 8 8 7	Districts, Blocks and Villages were selected purposively according to the majority of the tribes and according to the convenience of accessibility whereas respondent had been selected randomly
	Senapati (Total Blocks=6)	Paomata Tadubi	Tungjoy Kaibi Maikhel	Poumei Mao Kom	8 7 8	
	Tamenglong (Total Blocks=5)	Nungba	Kambiron	Kabui	8	
						With the Assistance of interpreter
Total	**3**	**5**	**7**	**9**	**71**	
	West Garo Hills (Total Blocks=6)	Rongram Gambigre	Chasingre Damalgre	Garo	4 4	Districts, Blocks and Villages were Selected purposively

ISBN: 978-93-85822-06-3

Meghalaya (11)	East Jaintia Hills (Total Blocks=2)	Khliehriat	Bapung	Jaintia	8	According to the majority of The tribes and according to the convenience of accessibility whereas respondent had been selected randomly with the assistance of interpreter Districts, Blocks and Villages were selected purposively according to the majority of the tribes and according to the convenience of accessibility whereas respondent had been selected randomly
	Ri-bhoi (Total Blocks=3)	Umsning	Umiam	Khasi	8	
Total	**3**	**4**	**4**	**3**	**24**	
Mizoram (8)	Champhai (Total Blocks=4)	Champhai	Kahrawt	Mizo	10	
Total	**1**	**1**	**1**	**1**	**10**	

ISBN: 978-93-85822-06-3

	Kohima (Total Blocks=7)	Kohima	Chedema Model village, Chadema village	Angami	8	Districts, blocks and villages were selected purposively according to the majority of the tribes and according to the convenience of accessibility whereas respondents had been selected randomly with the assistance of interpreter
				Ao	10	
				Sema	5	
		Tseminyu	Chunlikha, Gukhanyu,		4	
				Chakesang	4	
Nagaland (11)			Chiechama			
		Chiephobozou			8	
			Chiemekhuma	Rengma		
					5	
Sema						
	Wokha (Total Blocks=7)	Wokha	Changsu New, Changsu Old	Lotha	4	
					4	
			Yanmhon New,	Phom	4	
		Bhandari	Yanmhon Old		4	
Total	**2**	**5**	**10**	**8**	**60**	
Total:						
5	**12**	**19**	**30**	**25**	**225**	

ISBN: 978-93-85822-06-3

METHODS OF DATA COLLECTION

For collection of data, both interview and questionnaire methods had been followed.

Pretesting of Interview Schedule

Pre-testing is the process of an advanced testing of the study design after the schedule/questionnaire has been prepared. The objective of pre-testing is to detect the discrepancies that have crept in and to remove them after necessary modification in the schedule/questionnaire. The draft Interview Schedule was pre-tested in Umiam, Ri-bhoi district of Meghalaya.

Field Data Collection

Relevant data were collected with the help of schedule constructed for the study. In Manipur, data were collected during October to November, 2011, in Arunachal Pradesh it was collected during April, 2012, in Nagaland it was collected during May, in Mizoram it was collected during June and in Meghalaya it was collected during September, 2012. The data was collected by the researcher with the help of interpreter in some places.

Field Data Collection at a glance

States	Time of Data Collection
Arunachal Pradesh	April, 2012
Manipur	October-November, 2011
Meghalaya	September, 2012
Mizoram	June, 2012
Nagaland	May, 2012

Mail Questionnaires

In three of the twelve districts where a researcher finds difficulty in taking personal interview, mail questionnaire method was used where structured

ISBN: 978-93-85822-06-3

interview schedule was mailed to the respondents and after filling up the schedule, it was mailed back to the researcher. Thirty nine samples were taken through this method.

Secondary Data Collection

Secondary data relating to the demographic features of the state has been collected from published materials so far available from the State Agricultural Department, census reports, Directorate of Economics and Statistics of all the concerned states. Data related to the tribes are collected from available literature, books and also from the internet

Variables and their Measurements

Operational definitions of 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.

Occupation (X_2): Occupation of farm community refers to the engagement of the farm community for their major source of livelihood during the major part in the year. The scale has been ascribed according to the judge's ratings. The schedule developed for the study

Schedule developed for the study:

Category	Scale
Labour	1
Business	2
Cultivation	3
Service	4

ISBN: 978-93-85822-06-3

Education (X_3): 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 the respondent is studying

Family Size (X_4): It denotes the total no. of persons living in the respondents' house under a single household.

Social Participation (X_5): Operationally it has been defined as the degree to which the respondents were involved in formal organisation as member of or office bearers and regularity in their attendance to meeting. It has been estimated by multiplying the number of meetings they attended in a year with the number of organisations they are engaged with. The scale use for this present study was developed by Hay (1991).

Category	Scale
Not a member of organisation	0
Member of one organisation	1
Member of more than one organisation	2
Office holder	3
Wider public leader	4

Economic Status (X_6): It is the materials possession of an individual. The summation of the score of all five items divided by the number of items indicated the economic status of an individual farmer. For this present study, the following assessments developed by Pareek and Trivedi (1964) were used with a slight modification.

- **Land**: No land(0), less than 1 acre(1),1-5 acre(2), 5-10 acre(3),10-15 acre(4),15-20acre(5) more than 20 acre(6)
- **Hous**e: No home (0), Hut (1), Kutcha (2), Mixed (3), Pucca (4), Mansion (5)

ISBN: 978-93-85822-06-3

- **Farm Power:** No draught animal (0), 1-2 draught animal (1), 2-4 or more prestige animal (2), 5-6 draught animal or tractor or power tiller (3)
- **Mechanical possession:** Bullock cart (1), Radio (1), chair (1), Cycle (1), improved agricultural implement (2)/TV/Refrigerator/Car/ motorcycle (3)
- **Level of Sanitation:** Bad (1), medium (2), Good (3)

Outside Contact (X_7): Outside contact is the degree to which an individual is exposed beyond his or her periphery. Outside contact was measured by ascertaining from the respondents the frequency of places they visited and exposed for important requirements like the state capital, district town, sub division town and nearest town. The scale has been ascribed according to the judge's ratings. The schedule developed for the study:

Places	**Most Often (4)**	**Often (3)**	**Some Time (2)**	**Rarely (1)**	**Never (0)**
Aizawl/Shillong/Itanagar/Imphal/Kohima					
District Town					
Sub Division Town					
Nearest town					

Contact within Periphery (X_8): Contact within periphery is the degree to which an individual has a contact within his or her close circles. It could be isolated and localized. The data was taken by measuring how often do a respondent has a contact with the neighbours, family members, relatives,

ISBN: 978-93-85822-06-3

family friend, local leaders and friends for gathering information and basic needs. The scale has been ascribed according to the judge's ratings. The schedule developed for the study:

Source Neighbours	**Most Often (4)**	**Often (3)**	**Sometime (2)**	**Rarely (1)**	**Never (0)**
Family Members					
Friends					
Family Friend					
Relatives					
Local Leaders					

Mass Media Exposure (X_9): Mass media exposure is the degree to which an individual is exposed to the mass media with respect to new innovation in agriculture, economic activity and other information. Mass media exposure was therefore measured by ascertaining from the respondents about the frequency of watching television, listening to radio programmes, reading newspaper as well as other types of exposures. This was measured with the help of scale developed by Singh (1972).

Source	**Most Often (4)**	**Often (3)**	**Sometime (2)**	**Rarely (1)**	**Never (0)**

ISBN: 978-93-85822-06-3

Daily Newspaper Weekly Magazine					
Monthly Magazine					
Govt. Publication Political					
Literature					
Television					

Interpersonal Source (X_{10}): Interpersonal source was estimated by sources comprising of different localite and cosmopolite personnel which the respondents used as a channel to get information of different areas where the followings were included: members of different groups, village people, school teacher, B.D.O, A.D.O, and NGO's. The scale has been ascribed according to the judge's ratings.

The schedule developed for the study:

Source Members of the other Groups	Most of Times(3)	Some Times (2)	Less of Time (1)	Never (0)
Village people				
School Teacher				
B.D.O				

ISBN: 978-93-85822-06-3

A.D.O				
NGO/ Facilitator				

Interactive Information (X_{11}): Interactive information were estimated by accessing the frequency of interaction and the extent of usefulness with institutional sources such as banks, programme officials, rural institution and village level workers. The scale has been ascribed according to the judge's ratings. The schedule developed for the study:

Source	Most of Times(3)	Some Times (2)	Less of Time (1)	Never (0)
Members of the other Groups				
Village people				
School Teacher				
B.D.O				
A.D.O				
NGO/ Facilitator				

Capacity Information (X_{12}): Capacity information was conceived in the study as the kind of information that helps capacity building and generating impact in communication. Here in this study capacity information has been estimated from training, exhibition, radio, television, newspaper and magazine. The scale has been ascribed according to the judge's ratings. The schedule developed for the study:

ISBN: 978-93-85822-06-3

Source Training	Most Useful	Somewhat Useful	Less Useful
Exhibition			
Radio			
Television			
Newspaper/ Magazines			

Body Language (X_{13}): It comprises of the following gestures through which transmission of information and expression of emotions are organised. Scale has been developed for the present study:

- **Facial expression:** Angry (1), Moody (2), Normal (3), Smiley (4)
- **Posture:** Shaking Shoulder (1), Bowing forward (2), Straight without movement (3), Action (4)
- **Use of touch:** Touches (1), Do not touch (2)
- **Use of clothing (general):** Casual (1), Colourful (2), Accessories (3)
- **Use of clothing (special occasion):** Neutral (1), Colourful (2), Accessories (3)

Kinesics (X_{14}): It is basically a body language, but for this study, some unique body languages were measured as kinesics. The scale has been ascribed on the basis of judge's rating.

- **Tone of voice:** Loudly (1), Softly (2), Normal (3)
- **Speed of speaking:** Rapid (1), Medium (2), Slow (3)

ISBN: 978-93-85822-06-3

- **Use of eye:** Close (1), Blinking (2), Open (3)

Contingency Communication (X_{15}): It is a kind of communication which is being organised to meet up the contingent need. It has been estimated under the following items. On the basis of the judge's rating the value was ascribed.

- Information about the innovation: (agriculture/health/education): Radio (1), Local leaders (2), local newspapers (3), television (4)
- Communication in terms of emergency: Newspaper (1), Television (2), Telephone (3)

Communication Pattern in observing certain rituals (during festivals and marriages): Local newspaper (1), Local TV channel (2), announcement in church (3) invitation card (4)

Operational definitions of Dependent Variables:

Mandays (y_1): The operational definition of mandays here has been conceived in terms of number of men/women days generated by agriculture and others which contributed the income.

Level of Decency (y_2): Level of decency here in the study has been perceived as the degree to which an individual found his job decent or not. It was estimated by multiplying the score with their perception of decency taking a ten scale. The rating was ascribed according to the judges' ratings. The scale developed for this study was:

Enjoyable (1)/ beneficial (2)/ socially rewarding (3)/ graceful (4)

Livelihood (Y_1): Livelihood in this study has been estimated by summing up the mandays and level of decency.

ISBN: 978-93-85822-06-3

Security Perception (y_3): Security perception is the degree to which an individual perceived his job as well as his daily food intake as secure or not. The rating was ascribed according to the judges' ratings. The scale developed for this study was:

Free of hazards (4)/ risks (3)/danger (2)/ignominy (1)

Calorie Intake (y_4): Calorie Intake was estimated by the amount of calorie consumption of the respondents from their daily food intake. It was calculated by adding the calorie contents of different foods they consumed in a day.

Food Security (Y_2): Food Security in this study was estimated by adding the security perception and calorie intake.

Income from Agriculture (y_5): 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.

Income from Subsidiary (y_6): 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.

Expenditure on Health Care (y_7): It is conceived as the expenditure incur after health and its related concerns. It was calculated by dividing the yearly expenditure on health care divided by family members.

Economic Security (Y_3): Economic security was calculated by summing up income from agriculture, income from subsidiary and expenditure after health care.

ISBN: 978-93-85822-06-3

Agricultural Development (Y_4): Agricultural development was estimated by the degree to which an individual is exposed to new technology as well as the production and yield of the major crops he cultivated.

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.

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

Where, X=mean of the distribution

f= frequency of the class

x= class value or midpoint of the class interval

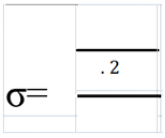
N= number of observations

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

ISBN: 978-93-85822-06-3

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:



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

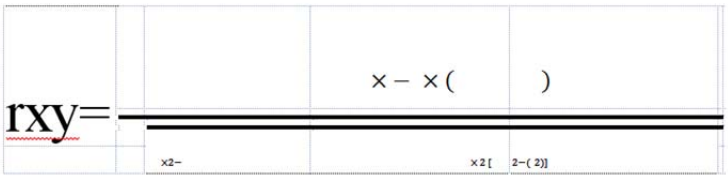
$$\text{Co-efficient of variance (C.V.)} = \frac{\quad}{\quad} \times 100$$

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

ISBN: 978-93-85822-06-3

ratio level of measurement and are linearly related. A person product moment (r) is computed by the formula:



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

N= number of paired scores

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

$\sum$×= sum of X score

$\sum$×^{2}=each X squared, then summed

($\sum$×)^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 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.

ISBN: 978-93-85822-06-3

Path Analysis

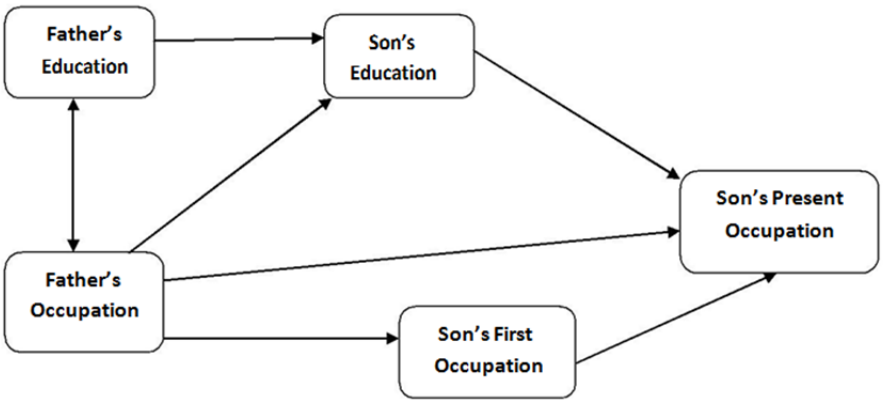
The term was first introduced by the biologist Sewal Wright in 1934 in connection with decomposing the total correlation between any two variables in a causal system. The technique is based on a service of multiple regression analysis with the added assumption of the causal relationship between independent 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 making explicit the assumption, underlying the causal connections and then by elucidation the direct effect the explanatory variables.

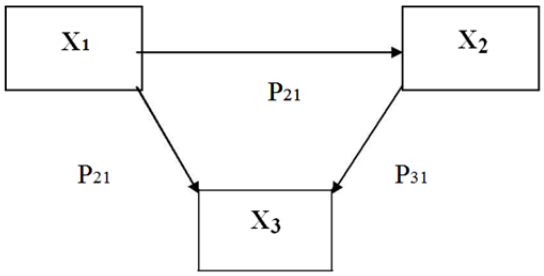
An illustrative path diagram showing inter relationship between father's education, father's occupation, son's first and son's present occupation can be shown as;

ISBN: 978-93-85822-06-3



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).

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)

ISBN: 978-93-85822-06-3

The structural equation for the above can be written as:

$X_1 = e_1$

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

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

X_1 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.

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

Multiple Regression Analysis

Multiple regression analysis technique was used to determine the degree to which the dependent variable could be predicted with the independent

ISBN: 978-93-85822-06-3

variables and also to trace out the contributory influence of independent variables on dependent variables. The following prediction equation was usd for this study to determine the partial regression coefficients:

$$Y=a+b_1x_1+b_2x_3+,\ldots\ldots+b_nn_n$$

Where,

Y=dependent variables

a=constant term

b1…bn=partial regression coefficients, and

n=number of independent variables

Each regression coefficient (b) which appears in the above equation represents the amount of change in Y that can be associated with a given change in any one of the X's with remaining independent variable held fixed.

The multiple regression tool as also used to compute coefficient of multiple determination (R2). This gives the percentage of variation explained by the independent variables (X1, X2,…Xn) in dependent variable (Y). The R2 can be calculated by using formula:

$$R2=\frac{\sum biS1Y}{SYY}$$

Where ,

R2 = Coefficient of multiple determination

ISBN: 978-93-85822-06-3

B1 = Partial regression coefficient of Y on X1

S1Y = ∑X1Y1, and

$$SYY = \sum Y2-\frac{(\sum Y)2}{n}$$

Factor Analysis

Factor analysis is a method for investigating whether a number of variables of interest Y1, Y2, …. Yl, are linearly related to a smaller number of unobservable factors F1, F2, ….. Fk . The fact that the factors are not observable disqualifies regression and other methods previously examined. However, that under certain conditions the hypothesized factor model has certain implications, and these implications in turn can be tested against the observations. Exactly what these conditions and implications are, and how the model can be tested, must be explained with some care.

Exploratory Factor Analysis (EFA): Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and on the linear relationships between the observed variables and the factors but only specifies the number of latent variables.

- Confirmatory Factor Analysis (CFA) Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population – characterized by allowing restrictions on the parameters of the model

ISBN: 978-93-85822-06-3

Applications of Factor Analysis

- Personality and cognition in psychology
- Child Behavior Checklist (CBCL)
- MMPI
- Attitudes in sociology, political science, etc.
- Achievement in education
- Diagnostic criteria in mental health

The Factor Analysis Model

The factor analysis model expresses the variation and covariation in a set of observed continuous variables y (j = 1 to p) as a function of factors η (k = 1 to m) and residuals ε (j = 1 to p). For person i,

$yi1 = v1 + \lambda 11\ \eta i1 + \lambda 12\ \eta i2 + \ldots + \lambda 1k\ \eta ik + \ldots + \lambda 1m\ \eta im + \varepsilon i1$

$yij = vj + \lambda j1\ \eta i1 + \lambda j2\ \eta i2 + \ldots + \lambda jk\ \eta ik + \ldots + \lambda jm\ \eta im + \varepsilon ij$

$yip = vp + \lambda p1\ \eta i1 + \lambda p2\ \eta i2 + \ldots + \lambda pk\ \eta ik + \ldots + \lambda pm\ \eta im + \varepsilon ip$

where

vj are intercepts

λjk are factor loadings ηik are factor values

εij are residuals with zero means and correlations of zero with the factors

In matrix form,

$\boldsymbol{y}\mathrm{i} = \boldsymbol{v} + \boldsymbol{\Lambda}\ \boldsymbol{\eta}\mathrm{i} + \boldsymbol{\varepsilon}\mathrm{i}$, where

$\boldsymbol{v}$ is the vector of intercepts vj,

$\boldsymbol{\Lambda}$ is the matrix of factor loadings λjk,

$\boldsymbol{\Psi}$ is the matrix of factor variances/covariances, and $\boldsymbol{\Theta}$ is the matrix of residual variances/covariances

ISBN: 978-93-85822-06-3

with the population covariance matrix of observed variables $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\ \boldsymbol{\Psi}\ \boldsymbol{\Lambda}\ +\boldsymbol{\Theta}.$$

Canonical Discriminant Function

Discriminant function analysis is used to determine which continuous variables discriminate between two or more naturally occurring groups. For example, a researcher may want to investigate which variables discriminate between fruits eaten by (1) primates, (2) birds, or (3) squirrels. For that purpose, the researcher could collect data on numerous fruit characteristics of those species eaten by each of the animal groups. Most fruits will naturally fall into one of the three categories. Discriminant analysis could then be used to determine which variables are the best predictors of whether a fruit will be eaten by birds, primates, or squirrels.

Discriminant analysis linear equation

DA 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:

$D= v_1 X_1+ v_2 X_2+ v_3X_3= .....v_i X_i + a$

Where D = discriminate function

v = the discriminant coeffi cient or weight for that variable

X = respondent's score for that variable

a = a constant

i = the number of predictor variables

ISBN: 978-93-85822-06-3

This function is similar to a regression equation or function. The v's are unstandardized discriminant coefficients analogous to the b's in the regression equation. These v's maximize the distance between the means of the criterion (dependent) variable. Standardized discriminant coefficients can also be used like beta weight in regression. Good predictors tend to have large weights. What you want 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 classifi ed. The number of discriminant functions is one less the number of groups. There is only one function for the basic two group discriminant analysis.

ISBN: 978-93-85822-06-3