



USAID | INDIA
FROM THE AMERICAN PEOPLE



Advance Analysis of Secondary Data Using Multivariate Statistical Techniques

Workshop Report

**September 29 – October 1, 2008
Conference Hall
NIHFW, New Delhi**

Submitted by:

MCH-STAR Initiative Prime Partners

Emerging Markets Group, Ltd.

Boston University

The Centre for Development and Population
Activities (CEDPA)





USAID | INDIA
FROM THE AMERICAN PEOPLE



Workshop on

**Advance Analysis of Secondary Data
Using Multivariate Statistical
Techniques**

September 29th – October 1st, 2008
**Conference Hall National Institute of Health and Family
Welfare(NIHFW)**



Contents

Brief Background.....	2
Objectives	2
Name(s) of facilitator(s).....	2
Sessions.....	3
MNCHN Indicators.....	3
Association and Test of Significance.....	3
Statistical Techniques: Correlation and Regression	3
Statistical Technique: Logistic Regression Analysis.....	3
Statistical Techniques: Introduction to MCA and Multilevel Analysis.....	3
Introduction to Factor Analysis	4
Understanding Survey Data with Special Reference to National Family Health Survey.....	4
Developing Conceptual Framework	4
Data Analysis on Identified Topics by Groups.....	4
Group Work to undertake Multivariate Analysis.....	4
Publication Processes.....	5
Group Presentations on Multivariate Findings – Discussions and Interpretation.....	5
Next Steps, Closing and Workshop Evaluation.....	5
Highlights.....	5
Challenges.....	5
Output	5
Summary of evaluation	5
Next Steps	6
Attachment 1: AGENDA.....	7



Workshop on Multivariate Analysis

Brief Background

MCH-STAR organized a series of workshops to build capacity of public health professionals on data analysis and its meaningful interpretation to improve Maternal Neonatal and Child Health and Nutrition (MNCHN) status in India. Towards this endeavor, the first workshop on 'Secondary Data Analysis' was conducted in the month of August in which officials from Star Supported Institutes (SSIs) actively participated and identified priority topic of research that needs to be analyzed further by applying multivariate and advance statistical techniques.



The second in this series of workshops, 'Advance Analysis of Secondary Data using Multivariate Analysis' was held during September 29 – October 1, 2008 at the National Institute of Health and Family Welfare (NIHFW), New Delhi.

The participants of this workshop were those who had already attended the Secondary Data Analysis Workshop. They worked on actual secondary data analysis on topics which were identified through stakeholder consultative meetings at various points in time in recent past.

The primary purpose of this workshop was to provide an orientation on advance statistical techniques (including multivariate analysis techniques) as well as providing hands-on-practice in analyzing secondary data.

Objectives

The specific objectives of the workshop are to:

1. Learn about various multivariate analysis techniques, data requirement and examples of their uses in the field of MNCHN,
2. Understand handling statistical packages to analyze data using various statistical techniques,
3. Hands on practice on analysing secondary data such as NFHS and DLHS using multivariate technique for deriving meaningful results on the identified topics, and
4. Understanding meaningful interpretation of results

Name(s) of facilitator(s)

Dr. Sulabha Parsuraman, IIPS, Mumbai
Dr. Pradeep Deshmuk, MGIMS, Wardha,
Dr. Damodar Sahu, NIMS, New Delhi
Dr. Sanjay Kumar, MCH-STAR, New Delhi
Dr. Ashok Patwari, MCH-STAR, New Delhi



Sessions

The workshop started with welcoming participants and providing a brief overview of the previous workshop on secondary data analysis, particularly the identified topics of research using secondary data and applying multivariate statistical techniques. Many of the participants were those who attended the previous workshop. The agenda and the facilitators of the workshop were introduced. There were three external facilitators for this workshop along with officials of MCH STAR.

MNCHN Indicators

The first session of the workshop was on MNCHN indicators in which participants were provided with an overview of characteristics of a good indicator (e.g. reliability, comparability, precise definition, consistent with international standards etc.). In addition discussion took place about the logical framework model of indicators by defining and sequencing input, process, output, outcome and impact indicators. The session ended with examples of indicators on MNCHN issues such as breastfeeding and nutrition, maternal and reproductive health, child health etc.

Association and Test of Significance

The second session outlined various measures of association and test of significance to consider while analyzing data. Some basic concepts of statistics such as population, sample, and parameter were discussed. This was followed by a number of basic biostatistics principles such as: steps for hypothesis testing; types of errors; properties of normal distribution; chi-square test of significance; and test of goodness of fit. This session was designed to present some of the basic tests of significance used to find association between variables.

Statistical Techniques: Correlation and Regression

As a first step of using advance statistical techniques, this session provided an overview of: correlation and regression analysis; type of variables needed for such analysis; and assumption underlying use of these techniques. During the presentation live examples of data and computation of correlation and regression coefficients were demonstrated. During this session concepts of partial correlation and multiple regression models were also discussed.

Statistical Technique: Logistic Regression Analysis

This session provided an overview of one of the most commonly used multivariate technique in the field of health and nutrition research- logistic regression. This method is most appropriately used when the variable is measured on categorical scale or is dichotomous. The presentation first defined logistic regression and then discussed how it provides you clear understanding of the effect of each predictor variables on dependent variable after controlling for other independent variables in the model. Furthermore, the presentation emphasized model requirement such as type of response and predictor variables. With the help of an example and tabulated results, the discussion focused on interpretation of the results of logistic regression.



Statistical Techniques: Introduction to MCA and Multilevel Analysis

Another multivariate statistical technique known as Multiple Classification Analysis (MCA) was discussed in this session. This technique is most appropriate when the dependent variable is on an interval scale and the independent variables are categorical or dichotomous. Another advantage of this method is the use of covariates which are known to affect the dependent



variable, however, the covariates needs to be measured on interval scale. An example of use of MCA was also presented and its interpretation was attempted during this session.

An introduction to multi level analysis was presented briefly during this session. This analysis uses data at micro level (individual respondents) and community variables (village level information) and combines them together to derive effects of variables measured at different levels in a model. Only the concept of multi level analysis was discussed without much detail. Participants we keen to participate in another workshop to understand this technique in more detail.

Introduction to Factor Analysis

The session on factor analysis discussed: the purpose of using this technique; methods of data reduction without losing information; example for using factor analysis; and the procedure to derive results from a statistical package. The outputs generated by the package were then discussed to help participants understand the tabulated results and how to interpret them.

Understanding Survey Data with Special Reference to National Family Health Survey

This session was devoted to explain various steps by which secondary data particularly from the National Family Health Survey could be obtained. A detailed description of these steps by going through web browsing was demonstrated. This was followed by presentation on the file structure of NFHS data including: household records; women's records; men's' records; and children's records. The presentation discussed the scope of the information available in the survey data. This session also discussed some of the basic steps before using raw data such as using proper weight to derive meaningful results.

Developing Conceptual Framework

The purpose of this session was to discuss the importance of a conceptual framework for investigating any research question. A presentation was made with an example of an important published conceptual framework 'An Analytical Framework for the Study of Child Survival in Developing Countries' by Mosley and Chen. The discussion centered around developing sequence of variables with an cause and effect relationship as a part of undertaking research which should also take care of confounding variables in establishing any relationship among dependent and independent variables.



This session was followed by group work to develop conceptual framework of data analysis for topics identified by each of the SSI in the last workshop. Presentations were given of the various conceptual frameworks and feedback was provided to improve the framework for each topic.

Data Analysis on Identified Topics by Groups

After getting inputs on the conceptual framework on each of the identified topics, the participants started handling data. With the support of the facilitators, SSI's selected and recoded variables to be used in accordance with their framework. Towards the end of the day, each group presented findings based on uni-variate and bi-variate analysis and received comments from the facilitators and other participants.

Group Work to undertake Multivariate Analysis



After undertaking univariate and bi-variate analysis, the group started applying multivariate statistical techniques and generated output tables. Some used STATA to handle and analyze NFHS 3 raw data while others used the SPSS package. This session and the previous session were geared at providing hands on training on data handling, analysis and interpretation of output tables generated by the software. This allows our SSIs to undertake similar analysis in future on any given topic.

Publication Processes

This session was designed to provide various steps, process and protocol for preparing and submitting any research article to reputed, peer reviewed journal. The presentation highlighted: how to choose an appropriate journals for your paper; authorship and criteria for granting authorship; copyright; plagiarism; competing interests; addressing comments provided by reviewers; and other tips for writing papers.

Group Presentations on Multivariate Findings – Discussions and Interpretation

Each group presented their initial findings on the analysis and use of multivariate techniques. A lively discussion took place towards interpretation of the results. The facilitators and participants gave their inputs on the presentation for further exploring data and firming up the results.

Next Steps, Closing and Workshop Evaluation

The workshop ended with identification of follow up activities towards preparation of paper on each of the identified topics. Each SSI discussed and provided a roadmap towards producing finer results. These road maps were submitted to MCH-STAR for comments so that a paper could be developed.

Highlights

The main highlight of the workshop was the hands on training of the participants on data handling, analyzing and deriving preliminary results on their respective topics. They were able to use multivariate techniques and interpret the results after undertaking the data analysis.

Challenges

Some SSI's representatives, like IndiaCLEN, put together the initial work on some great ideas in the workshop however, since participants are located in various cities, it is challenging for them to coordinate and collate results and develop paper out of them. Though an attempt was made to distribute specific roles and responsibilities, coordination among them remains a challenge. The Public Health Foundation of India representatives seemed distracted throughout the workshop. They were also not very open to the feedback given by the group on their presentation.

Output

The workshop witnessed analysis of raw data from NFHS 3 survey and preliminary findings on the identified topics. Other output of this workshop was building the capacity of the participants for use of secondary data with live example so that they could carry out more research in future.

Summary of evaluation

An evaluation form was filled in by the participants to express their level of satisfaction with each of the sessions. Overall majority of the participants rated various session as very good and good, while



some of the session it was felt too statistics driven especially for those who did not have statistics background.

Next Steps

As a follow up of this workshop, MCH-STAR will work with each SSI and see that the data are furthered analyzed and where appropriate research paper are produced. Depending upon the need, MCH-STAR will also undertake handholding for any assistance required by the SSIs in data analysis and use of multivariate techniques.



Attachment 1: AGENDA

**Advance Analysis of Secondary Data Using Multivariate Statistical Techniques
September 29 – October 1, 2008,
Venue: National Institute of Health and Family Welfare (NIHFW), NEW DELHI**

Time	Day 1 – Monday, 29 September 2008
9:00 – 9:15 am	Welcome, Introduction and objectives of the Workshop – <i>Dr. Marta Levitt-Dayal</i>
9:15 – 9:30 am	Recap on previous workshop on Secondary Data Analysis and Agenda of the present Workshop – <i>Dr. Sanjay Kumar</i>
9:30 – 10:00 am	MNCHN Indicators – <i>Dr. Damodar Sahu</i>
10:00 – 11:15 am	Association and Test of Significance – <i>Dr. Sulabha Parsuraman</i>
11:15 – 11:30 am	TEA
11:30 – 1:00 pm	Statistical Techniques : Correlation and Regression – <i>Dr. Pradeep Deshmukh</i> Group work : Handling data to compute Correlation and Regression coefficients in SPSS / STATA <i>(Group Work Facilitators: Dr. Sulabha Parsuraman, Dr. D. Sahu, Dr, Pradeep Deshmukh & Dr. Sanjay Kumar)</i>
1:00 – 2:00 pm	LUNCH
2:00 – 3:30 pm (Including Tea)	Statistical Technique : Logistic Regression Analysis – <i>Dr. Sulabha Parsuraman</i> Group work : Hands-on practice of undertaking Logistic Regression Analysis using SPSS / STATA <i>(Group Work Facilitators: Dr. Sulabha Parsuraman, Dr. D. Sahu, Dr, Pradeep Deshmukh & Dr. Sanjay Kumar)</i>
3:30 – 4:15	Group Work Presentations
4:15 – 5:45 pm	Statistical Techniques : Introduction to MCA and Multilevel Analysis – <i>Dr. Sulabha Parsuraman</i>



Time	Day 2 – Tuesday, 30 September 2008
9:00 – 9:15 am	Recap of previous day
9:15 – 9:45 am	Introduction to Factor Analysis – Dr. Pradeep Deshmukh
9:45 – 10:15 am	Understanding Survey Data with Special Reference to National Family Health Survey (NFHS – III) – Dr. Sanjay Kumar
10:15 – 10:30 am	Developing Conceptual Framework – Dr. Damodar Sahu
10:30 – 10:45 am	TEA
10:45 – 12:00 pm	Developing Conceptual Framework for Identified Topics of Study : Group Work – Dr. Damodar Sahu, Dr. Pradeep Deshmukh
12:00 – 1:00 pm	Data Handling (parallel session) Preparing Data File for Analysis through STATA – Dr. Damodar Sahu Preparing Data File for Analysis through SPSS – Dr. Pradeep Deshmukh and Sanjay Kumar
1:00 – 2:00 pm	LUNCH
2:00 – 3:15 pm	Group Presentations on Conceptual Framework and Discussions
3:15 – 3:30 pm	TEA
3:30 – 6:00 pm	Data Analysis on Identified Topics by Groups <i>(Group Work Facilitators: Dr, Pradeep Deshmukh, Dr. D. Sahu, & Dr. Sanjay Kumar)</i>



Time	Day 3 – Wednesday, 1 October 2008
9:00 – 9:15 am	Recap of previous day
9:15 – 10:30 am	Group Presentations on Findings from Univariate Analysis
10:30 – 10:45 am	TEA
10:45 – 1:00 pm	Group Work to undertake Multivariate Analysis <i>(Group Work Facilitators: Dr, Pradeep Deshmukh, Dr. D. Sahu, & Dr. Sanjay Kumar)</i>
1:00 – 2:00 pm	LUNCH
2:00 – 2:30 pm	Publication Processes – Dr. Ashok Patwari
2:30 – 4:00 pm	Group Presentations on Multivariate Findings – Discussions and Interpretation
4:00 – 4:15 pm	TEA
4:15 – 5:00 pm	Next Steps, Closing and Workshop Evaluation – Dr. Marta Levitt-Dayal

