



Proceedings of the Workshop
on
Large Scale Survey Data Management Using STATA
15th – 19th July 2024



Dr. Sandhya R Mahapatro
Coordinator

**Nabakrushna Choudhury Centre for Development
Studies, Bhubaneswar**

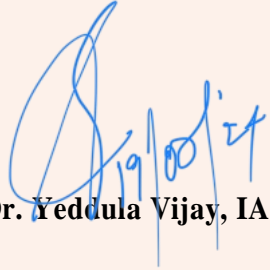


Message from the Director

The significance of quantitative research for evidence based planning is highly recognised among academicians and policy makers. Such type of research often relies on large scale data sets that offer insight into various socioeconomic issues of the society. Analysis of such data sets using various statistical and econometric techniques supports research and planning. In India various government organisations collecting large data sets in regular interval namely National Family Health Survey, National Sample Survey, Periodic Labour force Survey, India Human Development Survey, Annual Survey of Industries etc. Different statistical packages are applied for analysing and estimation of various economic, demographic and social indicators for the nation using such large scale data sets. STATA is one among the statistical software that enables the users to analyse, manage and visualise such data. Acquiring proficiency with the software aids in developing young researchers' capacity to produce research outputs with policy implications.

Given its immense significance, the Nabakrushna Choudhary Centre for Development Studies has organized a five day workshop on large scale socioeconomic data management using STATA from 15th to 19th July. The workshop has the twin objective of making participants understand large scale survey data and to familiarize them various statistical and econometric analysis through STATA software.

I am sure, the sessions taken and teaching materials shared by the resource persons who have come from different states of India was useful and help the researchers for their capacity building in this field. Finally, I would like to thank the organizing team including Dr Sandhya Mahapatro, the course coordinator for their tremendous effort and successfully conducting the much needed workshop.



Dr. Yeddula Vijay, IAS



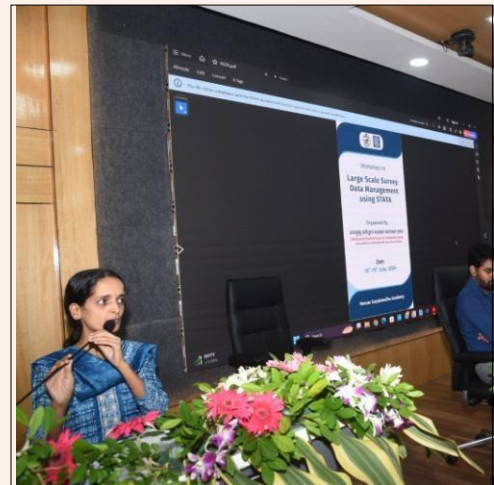
Introduction

In recent years use of quantitative data for social research and planning purpose is widely recognised. The large-scale survey data such as NSO, NFHS, and IHDS, collected on various development parameters give direction in the decision-making process for mobilising resources, planning and monitoring the programmes appropriately. For drawing meaningful results and informed policy decisions to improve livelihood, human capital, health of the population collection of such raw data required proper analyses and interpretation. The application of statistical package and use of econometric techniques using large-scale survey data will help the researchers and practitioners to address emerging challenges in research and come out with meaningful results for planning and programme. Among the different statistical software STATA is gaining more popularity and researchers familiarise with it would have the advantage of summarizing and managing data than those who have not that skill.

Against this backdrop, Nabakrushna Choudhury Centre for Development Studies, Bhubaneswar, an ICSSR institute in collaboration with Govt. of Odisha, organized a five-day workshop on large-scale socioeconomic survey data using the application of STATA software. The objective of the workshop was to provide a deeper understanding of large-scale data and its use. Hands-on training was provided for a better understanding of the data and statistical analysis. This helped capacity building of young researchers to come out with research outputs of policy significance.

The objectives of the workshop were:

- Familiarizing Large Scale Survey Data
- Extraction of NSS of data
- Introduction to STATA and its functions
- Coding, decoding, merging data sets
- Basic Statistical analysis using STATA
- Analysis of Quantitative Variables I
- Econometric modelling using STATA



Inauguration Ceremony



The workshop commenced with a welcome from Dr. Yeddula Vijay, the Director of NCDS. He underscored the indispensable role of data, not only in the field of economics but also in critical fields such as medicine and social



studies. Dr. Vijay lauded the workshop's objective of delivering hands-on training in STATA, emphasizing its relevance and timeliness. He expressed profound gratitude to the Department of Planning and Convergence, Government of Odisha, for their pivotal role in organizing the event. Dr. Vijay anticipated a rich and fruitful cross-learning experience, fostering collaboration between the distinguished resource persons and enthusiastic participants. This workshop heralds a significant revival of NCDS's proud tradition of hosting such intellectually enriching events, breaking a five-year hiatus and setting a promising

precedent for future endeavours.

Dr. Sandhya Mahapatro, Associate Professor at NCDS and the workshop organizer, eloquently emphasized the critical role of large-scale data in contemporary social science research. She highlighted how this workshop, breaking a five-year hiatus, serves as a pivotal platform for introducing emerging researchers to extensive datasets such as NFHS, NSSO, and PLFS. Dr. Mahapatro underscored the workshop's dual purpose: not only to familiarize participants with these essential surveys but also to equip them with the skills and capacity needed to translate these datasets into valuable insights, thereby enhancing their research endeavours and setting a promising precedent for future academic initiatives.

The keynote address was delivered by Shri Bijaya Bhushan Nanda, the Director of the Directorate of Economics & Statistics. Drawing from his extensive experiences at NCDS, he eloquently emphasized the crucial need for fostering a robust research culture to tackle pressing challenges being faced by the



society. Nanda delved into the multifaceted challenges inherent in social science research, highlighting the imperative for objective inquiry and the pivotal role of understanding and managing large-scale survey data.

He described the analytical process using STATA as "romancing with the data," capturing the intricate yet rewarding nature of data analysis. Nanda recognized the workshop as an essential stepping stone for participants aspiring to master these vital skills, setting the stage for their future contributions to the field.

The Chief Guest of the workshop, Shri G. Mathi Vathanam, IAS Director General, Gopbandhu Academy of Administration, conveyed his enthusiasm for the event and the opportunity to engage with young scholars. He shared his personal journey from engineering to the IAS, driven by a profound desire to focus on societal challenges and make a meaningful impact. Shri Vathanam highlighted several significant interventions in Odisha's largest slum, including a slum-development program that earned a prestigious UN Habitat award. He emphasized the critical importance of fieldwork in research, advocating for solutions-oriented studies that extend beyond theoretical frameworks to address real-world challenges effectively.

The session concluded with a vote of thanks delivered by Dr. Raj Kumar Khosla, Associate Professor at NCDS. He expressed deep gratitude to Shri Bijaya Bhushan Nanda, the keynote speaker, Dr. Yeddula Vijay, IAS the Director of NCDS, Shri G. Mathi Vathanam, IAS Director General, Gopbandhu Academy of Administration, and all the participants for their invaluable contributions. Special recognition was extended to Dr. Sandhya Mahapatro for her exceptional efforts in orchestrating the event, ensuring its success, and facilitating a seamless exchange of knowledge.



Workshop

The workshop commenced with an insightful session led by Dr. Subrata Mukherjee from IDS Kolkata. Dr. Mukherjee introduced participants to fundamental theories and concepts essential for undertaking statistical analysis. He elaborated on the growing significance of quantitative methods in social science research, highlighting how quantitative studies complement qualitative research to provide a more comprehensive understanding of social phenomena.

The second half of day 1 of the workshop was led by Dr. Mukesh Ranjan from Mizoram University. The session focused on the basics of STATA, including an introduction to the STATA manual, solving problems, renaming variables, data editing, unique ID generation, merging datasets, and appending data. The bare necessities that one has to master if they want to use large scale data surveys for their studies.

On the first half of Day 2 of the workshop, Dr. Ranjan provided a comprehensive overview of basic analysis techniques and explained essential statistical concepts necessary for data analysis. Participants were actively engaged in hands-on activities, practicing these techniques on sample datasets provided by Dr. Ranjan. The interactive nature of the session fostered an environment where participants could ask questions and receive immediate feedback, ensuring a solid understanding of the fundamental aspects of STATA. This practical approach not only reinforced learning but also boosted the confidence of the participants in handling large-scale survey data using STATA.



On the third day, Dr. Balakrushna Padhi from BITS Pilani led an insightful session centred on the extraction and analysis of NSSO data, with a particular focus on the Consumer Expenditure Unit Level Survey data. He elucidated the critical role of NSSO data in shaping policy at both state and central government levels, addressing pressing





issues such as income inequality, energy consumption, and poverty prevalence in India.



Dr.

Padhi delved into various methodologies, including the Modified Mixed Reference Period (MMRP), and emphasized the importance of different recall periods, such as the Uniform Reference Period (URP) and the

Mixed Reference Period (MRP), as outlined by prominent committees like those led by Lakhadawala, Tendulkar, and Rangarajan. He showcased the latest consumption expenditure data, discussing its implications for understanding regional disparities and inequality. Participants gained valuable insights into the practical applications of this data for policy analysis and research, particularly in the realms of education, migration, and expenditure.

In addition, Dr. Padhi conducted a comprehensive training session on extracting NSSO data using tools such as Nesstar and the dictionary approach in STATA, equipping attendees with essential skills for data manipulation and analysis. He also raptured the participants by teaching them how to draw various statistical graphs using STATA.

On the fourth day, Dr. JajatiKeshri Parida from Hyderabad Central University led an engaging session focused on the extraction and analysis of PLFS (Periodic Labour Force Survey) data. He provided practical, hands-on training on how to extract PLFS data, merge datasets, and perform preliminary analyses. Participants acquired valuable experience in calculating employment statistics, which enhanced their ability to conduct comprehensive labor





market analysis.

Dr. Parida's session was notably interactive, fostering active participant engagement in data manipulation and analysis tasks. This hands-on approach facilitated a deeper understanding of the intricacies of PLFS data and its application in research. His witty and humorous training style further enriched the learning experience, making the acquisition of these essential skills both enjoyable and effective.

Dr. Parida's session was notably interactive, fostering active participant engagement in data manipulation and analysis tasks. This hands-on approach facilitated a deeper understanding of the intricacies of PLFS data and its application in research. His witty and humorous training style further enriched the learning experience, making the acquisition of these essential skills both enjoyable and effective.





On the day 5 of the workshop, Dr. J.K. Parida delivered an enlightening session on regression modeling, delving into several crucial topics. Participants gained insights into Linear Probability Models (LPM), Logit, and Probit models, with a focus on interpreting STATA outputs for each model. The session provided strategies to tackle endogeneity issues, emphasizing the use of Instrumental Variable Probit (IV Probit) regression. Dr. Parida discussed methods to address self-selection problems, highlighting the Tobit and Heckman selection models. The fundamentals of regression modeling for time series and panel data were explored, including autocorrelation tests and the implementation of fixed effect and random effect models.

In the last session, Dr. Shakeel Ahmed led a session on introducing the DASP Package. Dr. Ahmed introduced the DASP package, offering detailed guidance on the complex installation process in STATA.

He underscored the package's importance in measuring inequality and poverty, focusing particularly on its use in estimating inequality within STATA. An overview of various inequality measures was provided, including the Gini coefficient, Atkinson index, and concentration index. Dr. Ahmed elaborated on the significance and limitations of these measures, giving participants a well-rounded understanding of their application and interpretation.



Valedictory Session

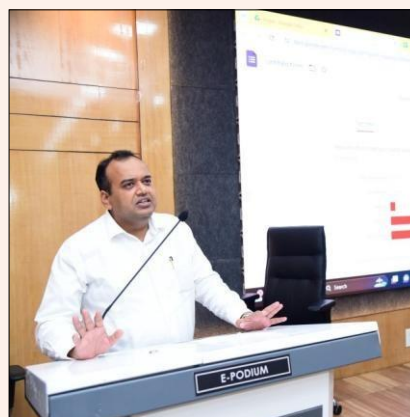
The valedictory session of the workshop commenced with Dr. Sandhya Mahapatro expressing her gratitude. She provided a succinct summary of the workshop, highlighting the key discussions and outcomes.

The chief guest, Shri Manish Aggarwal, IAS, Additional Secretary of the Department of Planning and Convergence and Additional Director General at the Gopbandhu Academy of Administration, underscored the indispensability of data in contemporary research. He emphasized its crucial role in addressing societal challenges and encouraged NCDS to organize more workshops to train young researchers in cutting-edge tools such as artificial intelligence, machine learning, and predictive analysis.

Dr. Vijay Yeddula, Director of NCDS, commended Dr. Mahapatro for her exceptional efforts in organizing the workshop. He concurred with Shri Aggarwal's remarks and committed to hosting more such workshops in the future. Dr. Yeddula also welcomed feedback from participants to enhance future events.

The feedback from participants was overwhelmingly positive. They expressed that the workshop was immensely educational, covering a broad array of topics in great depth. Many participants wished the workshop had been longer than five days to allow more time to delve into the subjects discussed.

Dr. Mahapatro extended her gratitude to Dr. Yeddula, her colleagues, and the entire NCDS staff for making the workshop a success. She offered special thanks to Mayadhar Shetty and Sneha Shrivastava for their extraordinary efforts, acknowledging their invaluable contributions. The workshop concluded with participants presenting Dr. Mahapatro with a memento, honouring her dedication and hard work





ସାମାଜିକ ବିକାଶ କ୍ଷେତ୍ରରେ ଓଡ଼ିଶା ସ୍ୱତନ୍ତ୍ର ପରିଚୟ ସୃଷ୍ଟି କରିଛି

ଭୁବନେଶ୍ୱର, ୧୫.୦୭ (ବନ୍ଦନା ସମ୍ଭାଷଣ)

ନବକୃଷ୍ଣ ଚୌଧୁରୀ ଗବେଷଣା କେନ୍ଦ୍ର ପକ୍ଷରୁ 'ଭାରତୀୟ ସମ୍ପଦା ସମାପନ ଯୋଜନାରେ ସାମାଜିକ ବିକାଶ କ୍ଷେତ୍ରରେ ସାମ୍ବଲ ଉତ୍ପାଦନ' ଉପରେ ଏକ ପାଠ୍ୟପୁସ୍ତକ ପ୍ରକାଶନ କାର୍ଯ୍ୟକ୍ରମ ଆୟୋଜନ କରାଯାଇଥିଲା।



ସମାଜିକ ବିକାଶ କ୍ଷେତ୍ରରେ ଓଡ଼ିଶା କେବଳ ବିକାଶରେ ଏକ ନିର୍ଦ୍ଦେଶାଳିତ ସ୍ୱତନ୍ତ୍ର ପରିଚୟ ସୃଷ୍ଟି କରିପାରିଛି। ସାମାଜିକ ବିକାଶ କ୍ଷେତ୍ରରେ ଆଧୁନିକ ଚର୍ଚ୍ଚା

ଅଗ୍ରଣୀ ଭାବେ ରାଜ୍ୟ ଅର୍ଥନୀତି ଓ ପରିଚାଳନା ବିଭାଗ ନିର୍ଦ୍ଦେଶକ ବିଜୟ କୁମାର ନାୟକ ଯୋଗଦେଇ ସାମ୍ବଲ ଉତ୍ପାଦନ ମହତ୍ତ୍ୱ ଉପରେ ଆଲୋଚନା କରିଥିଲେ। କାର୍ଯ୍ୟକ୍ରମରେ ନବକୃଷ୍ଣ ଚୌଧୁରୀ ଗବେଷଣା କେନ୍ଦ୍ର ନିର୍ଦ୍ଦେଶକ ଡ. ଯେନୁକା ବିହାରୀ ସାଗର ଦାଶଙ୍କୁ ପ୍ରମୁଖ ଭୂମିକା ସମ୍ଭାଳିଥିଲେ। ସମସ୍ତଙ୍କୁ ସମ୍ମାନିତ କରି ଉପସ୍ଥାପନା କରାଯାଇଥିଲା।

GOVERNMENT OF ODISHA | OFFICE OF THE SR. GENERAL MANAGER (HISTORICAL)





ORGANISING COMMITTEE

Dr Sandhya R. Mahapatro
Dr Raj Kumar Khosla
Dr Biswabas Patra
Mr Mayadhar Sethy
Mr Sneh Shrivastava
Mr Atul Kumar
Mr Sumanta Kumar Nayak
Ms Anuradha
Mrs Sumitra Pradhan



Nabakrushna Choudhury Centre for Development Studies
Bhubaneswar, Odisha-751013
Phone: 0674 - 2301094, 2300471, 2302602
e-mail: ncdsbbsr1987@gmail.com
<https://ncds.nic.in/>