



Param Poojya Dr. Babasaheb Ambedkar Smarak Samiti's
Dr. Ambedkar College Deekshabhoomi, Nagpur

One Day International Webinar on

“Large Scale Sample Surveys and Data Analysis”

Organized by

**Department of Statistics & Department of Computer Science
in association with Field Operations Division, National Statistical Office, Nagpur**

Date: 29th June 2020

REPORT

One day International Webinar on **“Large Scale Sample Surveys and Data Analysis”** was organized by Department of Statistics & Department of Computer Science in association with Field Operations Division, National Statistical Office, Nagpur on the occasion of 14th National Statistics Day to celebrate 128th birth anniversary of Padma Vibhushan Prof P.C. Mahalanobis. The main objective of this webinar was to disperse education to students all over India and abroad in these troubled times of crisis of COVID-19. This webinar was planned to impart knowledge about sampling surveys, different sample survey techniques, uses of collected data, analysis of collected data and coverage of data analysis for National Planning Commission. Webinar in particular highlighted the use of collection and analysis of data related to socio-economic aspects for government and for commercial use as well. It also broadly focused on how data analysis is used for drawing different timely and relevant conclusions which further help in decision making.

The keynote speaker for this webinar was Dr. Rajeeva Karandikar, Ex Director, Chennai Mathematical Institute and renowned psephologist. The three eminent speakers for this webinar were Shri. Ashok Toprani, Addl. Director General (Retd.), N.S.O., FOD, New Delhi; Shri. Rajant Meshram, Senior Manager, YouTube Creator, Google, Singapore; Shri. Rajendra Gautam, Dy. Director General, N.S.O., F.O.D., Nagpur Zone. Webinar was hosted by Mrs. A. M. Badar, Asso. Prof. of Dept. of Statistics and organizing secretary of the webinar,

Prof. R.V. Patil, Head of Dept. of Statistics and convenor of the webinar presented a welcome address. In the beginning Shri Patil Sir welcomed all the guest speakers, Principal of the college, staff members and the participants. He notified that this webinar was conducted keeping in mind the ‘14th National Statistics Day’, the day which is marked on birthday of Father of Statistics Prof P.C. Mahalanobis. He indicated that it is 128th birthday of Prof P.C. Mahalanobis. He gave information on immense contribution of Prof. P.C. Mahalanobis in the field of statistics and also how with the application of his statistical knowledge he made a remarkable nation building activity. He also explained how Prof Mahalanobis utilized his knowledge of statistics in the field of agriculture to produce home grown grains and helped in supply of the same in the country among his fellow citizens during shortage of food grains. Due to his immense contribution in statistics and in designing large scale sample surveys, Prof. Mahalanobis is being considered as Father of Modern Statistics of India. With this brief introduction about Prof. P.C. Mahalanobis and about the webinar Mr. R. V. Patil concluded his welcome note.

Principal of college Dr. Mrs. P. M. Siriya welcomed all speakers and everyone else on behalf of college and gave a brief introduction of college. She also congratulated all the organizing committee members of the Webinar.

Keynote Address by Dr. Rajeeva Karandikar.



Webinar began with inaugural address of Dr. Rajeeva Karandikar, Psychologist and Ex Director of Chennai Mathematical Institute, Chennai. In the beginning he mentioned that he fondly remembers his visit to Dr. Ambedkar college in the year 2011 for National Conference of Dept. of Statistics. He added to the earlier introduction given by convenor of webinar Shri. R. V. Patil of Prof. P.C. Mahalanobis that perhaps India was the only country where the central cabinet had statistical advisor before scientific advisor and this was essential for newly developing country and Prof Mahalanobis played important role as Statistical Advisor. He helped India in development through planning.

According to him sample survey or sample theory is that part of Statistics which requires least mathematical sophistication and therefore brighter students do not take interest in it but this is one part of Statistics which has wider applications with simple ideas but more impact. Statistics is not just a theory it is a real world connect and sample theory is important aspect of it. According to Dr. Karandikar students learn theory but they do not internalize it. He urged specially to students that they must internalize it.

According to Dr. Karandikar, Statistics is a tool which has both advantages and disadvantages. It can be misused as every tool can. He gave his favorite example of a knife. Knife in the hand of a surgeon saves life and in the hand of dacoit takes life. Knife is neither good nor bad, it is an enabler. Similarly, Statistics is an enabler. It is up to statistician how to use it. This is something that statistics students should internalize. Hence, he suggested that in any discussion if someone ridicules then students can come up with this defense.

Dr. Karandikar shared his experience with audience of which he was active part and the example was of Electronic Voting Machine (EVM) and its cross verification. The example was of last year Lok Sabha election. Prior to which there was huge controversy about electronic voting machine and cross verification. The people indicated them that at least 10% sampling be done because they think that a sampling fraction or sampling percentage determines an accuracy of cross verification of EVM votes but statisticians know that it is sampling size and not sampling fraction which determines accuracy of results.

Election commission constituted a committee with Dr. Karandikar as one of the members. A report submitted by the committee was countered and opposed by many officers and all opposition parties. All opposition party's discussion ended with demand of 50% sampling but fortunately Supreme Court did not agree with it. In case supreme court had agreed to it then expenditure of the order of 5 to 7 thousand crores on electronic voting machine and other things would have been wasted. Sample size defines the accuracy and not the sampling fraction which was not understood by laymen. He again emphasized that sample size defines the accuracy and not the sampling fraction and this was not understood by laymen.

He also mentioned about the use of two stage survey in crop cutting experiment done in West Bengal famine by Prof.P.C. Mahalanobis and idea of sequential testing has its origin in Mahalanobis work related to two stage sampling. Speaker stated that Statistics is a beautiful theory originated by practical application and great thinkers like Prof. Mahalanobis, C. R. Rao, R. A. Fisher, Karl Pearson, Neyman has given thought to concrete problems, built an abstract version and thus laid the foundation of statistical theory. He said while learning theory we need to focus on its deep-rooted applications. No research is needed in applied area to look at or solve a concrete problem. He mentioned that he used simple statistical ideas for solving concrete social problems of our country. He concluded by giving an advice to students to focus mainly on application of statistics, internalize the same and connect it with concrete problems.

Session-I “History of Large-scale Sample Surveys” by Shri. Ashok Toprani, Addl. Director General (Retd.), N.S.O., New Delhi. In the beginning he appreciated the organizers for conducting this webinar on National Statistics Day. He commenced by giving a brief introduction about National Sample Survey Office (N.S.S.O). He stated that Prof. Mahalanobis is father of Indian Statistical Systems in particular Large-Scale Sample Surveys. Since Speaker Mr. Ashok Toprani was a part of N.S.S. O for long time, he spoke about N.S.S.O. in detail. He informed that N.S.S.O is a premier government organization associated with large scale sample surveys in India since 1950. NSSO helped in planning of surveys, policy formulation and decision making.



Before starting his topic Mr. Toprani in support of previous speaker Mr. Karandikar’s point indicated that sample size is inefficient to draw conclusion at district or substate level as sample size is small to draw conclusions or to estimate any parameter. Then he stated that N.S.S.O uses Multistage stratified design for conducting survey. First stratification is done and then homogenous units are drawn. So even if any one unit is studied it helps in drawing a conclusion.

Mr. Toprani then started his session stating the flow of his address. He very well covered topics like divisions of N.S.S.O, users of the data collected by N.S.S.O., challenges faced by N.S.S.O and how those challenges are to be faced. He began to explain **4 divisions of N.S.S.O** which conduct large scale sample survey.

Survey Design and Research Division: This division designs and selects design for conducting surveys. It goes into multi stage stratified design. Sample size, Instruments, schedule, defining definitions and concept based on objective of survey are defined by this division. If required pre testing is also done in the field.

Field Operations Division: Speaker operated in this division. This division has regional offices across the country. These offices collect data from rural and urban parts of country and carry out surveys on different topics. Training about collecting data is given by zonal offices of this division.

Data Processing Center presently known as Data Quality Assurance Division: After data collection data is processed, pre data scrutiny, post data scrutiny, validations and generating tables required for final report is done by this division.

Survey coordination: This division coordinates between above 3 divisions of N.S.S.O. Central sample and state sample both on equal matching basis, conducts surveys and is coordinated by this division by using same instruments and definitions.

Speaker indicated that N.S.S.O. covers various topics of government interest and conducts surveys in a systematic way within set timeframe and maintains precision in collecting and processing information.

Then speaker specified **Users of N.S.S.O. data**

- Planning Commission of India known now known as Neeti Aayog for its plans.
- Central Statistical Organization.
- National Accounts Division (NAD)
- Government ministries.
- Reserve Bank of India (RBI)
- Other organizations.
- Researchers for their findings, interests and their decision-making process.

Mr. Toprani specified few topics on which surveys are conducted like Social Surveys in which surveys done till now are Consumer Expenditure, Employment, Unemployment, Education, Housing etc. Such surveys are carried out in housing area.

Economic surveys are carried out where there is data gap. Examples stated are unorganized manufacturing sector, Trade, Hotel, Restaurant etc. These surveys are carried out in establishments.

Mr. Toprani took his session forward by speaking about the **challenges**. Some of the challenges faced by N.S.S.O. are increased number of surveys, there is increased pressure of sample size as surveys are now taking place by bottom to up approach that is from district, state to country level. There are resource constraints as well. Sometimes focus on subject is lost due to detailed surveys. Contractual manpower, high alteration rate, training and retraining is also a challenge.

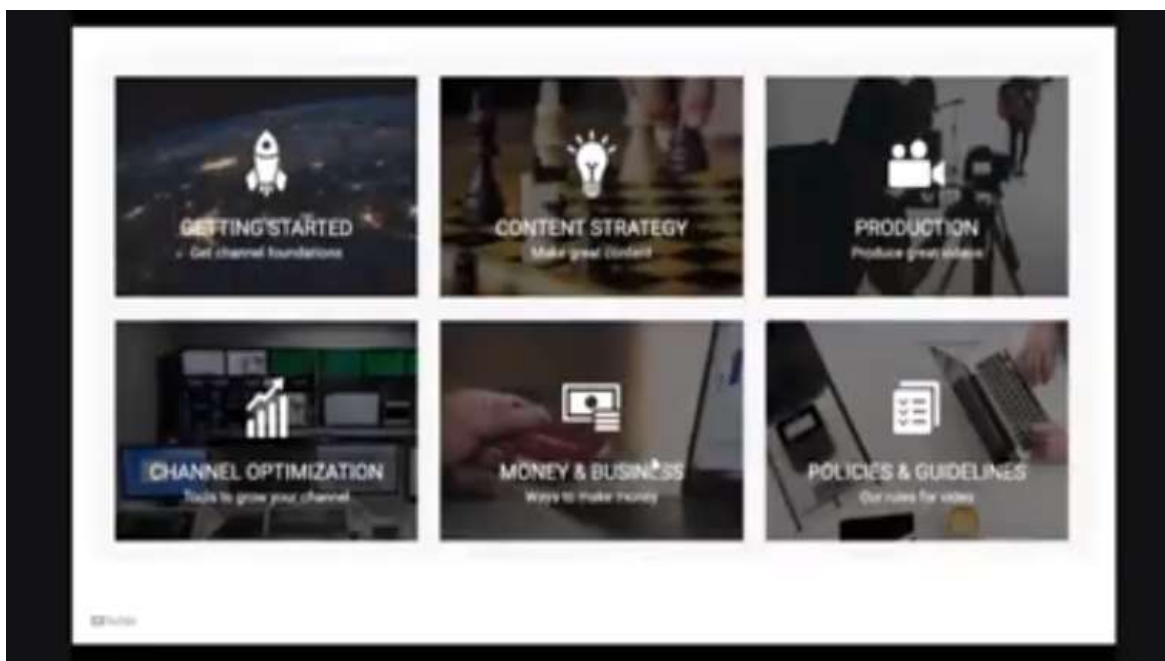
Mr. Toprani concluded his session by stating some of the **solutions** used by N.S.S.O. to overcome these challenges. Few stated solutions by Mr. Toprani being that shortening length of schedule as people don't have time for giving detailed and lengthy information. Interviewers' training is important so is the respondents time. If time is reduced more than the current time taken, a more appropriate information can be retrieved. Thus, in turn quality of response improves and hence findings, conclusions and decision making also improves. Questionnaire pattern is used for collecting information and this is working very well with regular manpower, contractual manpower and respondent. Thus, time taken for survey and reporting is minimized. He concluded his Session by conveying that how technology, which was once a challenge, helped to collect data quickly and precisely. Recently used interfaces are Computer aided telephonic interview, computer aided web interface and validation program for data validation. On his last note he urged the participants and students that they have a huge scope of research for improving the challenges for which they should focus on it.

Shri Toprani Sir's lecture was well appreciated by all participants and motivated the students to take more interest in Applied Statistics.

Session-II "Get Insights with YouTube Data Analytics" by Shri. Rajant Meshram, Senior Manager, YouTube Creator, Google, Singapore. He started his session with a presentation about YouTube Analytics and provided a link containing all the information that he presented in his session. Following is the link for the same.

<https://creatoracademy.youtube.com/page/home>

Speaker had a good audience interaction which he achieved by asking few general questions about YouTube. His session and presentation gave a detailed view of how YouTube Analytics can be used to make money on YouTube, how to earn audience for Your channel and how to enhance your video quality and content to get it more featured on YouTube. His session started by explaining



1) Understanding Video reach on YouTube

While explaining about this section he introduced two terms used in Youtube Analytics Impressions: When a viewer comes across certain video thumbnails on YouTube, it's called **impressions**, that is, how many times certain video thumbnails are shown on YouTube. Think of each impression as potential reach on YouTube and an opportunity to earn a view. An impression is counted right away if a viewer clicks on the thumbnail. Clicking on a link does not count as an impression. **Spikes in impressions can be common soon after upload. It's common for impressions to decline after the initial spike as the number reaches a steady state.**

CTR: Is often highest right after upload when some of most passionate fans and subscribers of a certain video might see the video at the top of their home page and subscribers' feeds. As total impressions grow beyond certain core audience, CTRs are expected to decline until reaching a relatively stable percentage (this is totally normal).

Next Mr Rajant explained about how YouTube analytics are used to make money on YouTube.

2) How much money one can make on Youtube

A channel to earn money must have 1,000 subscribers and 4,000 watch hours in the previous 12 month period in order to be eligible to apply to the YouTube Partner Program. Speaker indicated that content of the video uploaded must be meaningful and should be such that it can be watched by all age groups. Mr. Rajant explained the YouTube channel owner earns by enabling channel to different types of Ads and thus the money earned by this are shared by channel owner and Google/YouTube. Thus, types of Ads are:

Skippable Ads: Ads which can be skipped after 5 sec

Non-skippable Ads: Ads which cannot be skip.

Video discover Ad: these Ads describes the product.

Masthead: These are first to watch Ads on YouTube home page placed at the top section of YouTube Home page. These Ads are mostly used for big campaigns

YouTube helps to analyse which Ads are generating revenue for a certain channel and for which geographic location in channels Analytics tab.

Final topic Mr. Rajant focused on was

3) Who all are watching a certain YouTube channel

Speaker explained that a YouTube channel owner can login in YouTube studio and there is a tab known as Analytics which highlights highest views and lowest views so one can focus on increasing his viewers. YouTube analytics even helps to show which of the videos uploaded by a certain channel's owner are working and generating more revenue for them. Owner of channel can also compare two of his videos to focus on enhancing revenue.

Mr Rajant then explained viewers find a certain video by sending a link of a video/Viewers search in YouTube for their choice video/Browser featured/Viewer browse on app/Suggested video

Speaker concluded his session by checking knowledge of participants by asking few questions related to the topic explained.

Mr. Rajant had made his session very lively by involving participants by asking M. C. Q. related to hie topic. His presentation was very well appreciated by the participants.

Session-III In the beginning Shri. Rajendra Gautam, Dy. Director General, N.S.O., F.O.D., Nagpur Zone started his session on "Statistical Systems in India" by paying respect to Prof. P. C. Mahanalobis. Mr. Rajendra highlighted the great contributions of Prof. P.C. Mahanalobis towards country. He also congratulated everyone on 'National Statistics Day' and also stated that his office N.S.O Nagpur celebrates this day by organizing various competitions.



Mr. Rajendra started his session by a presentation on 'Statistical Systems in India' by N.S.O Nagpur. He covered topics like background of statistical function undertaken by various government agencies and elaborated on different statistical activities undertaken by statistical departments associated with N.S.O.

He indicated that sample survey is an integral part of system and the division of administrative functions between Centre and State governments based on subject classification under following three lists 1) Union List 2) State List 3) Concurrent List. Statistics system decentralized laterally among central and decentralized vertically between union territories and state.

To ease the work of NSO office it is reorganized in three sections mainly

- 1) DG(Statistics): This section has different sub division like
National Accounts Division (NAD)
Economic Statistics division (ESD)
Price Statistics Division (PSD)
Social Statistics Division (SSD)
- 2) DGNSS known as Director General National Statistical Section. This section various other divisions as
SDRD Sample Design and Resource Design
DQAD Data Quality Assurance Division
FOD Field Operations Division
- 3) DG (Coordination, Administration & Policy). This section includes Coordination, Training and Vigilance.

Speaker in his next section "Mandate of Ministry" stated that Ministry of Statistics and Program Implementation is responsible for development of Statistical Systems of our country. It co-ordinates between Statistical work of Centre and State, advises other departments for Statistical Methodologies and Analysis, prepares estimates of National Accounts, distributes information among different agencies. Speaker elaborated here that ministry carries out different large-scale surveys like Socio -economic surveys, Urban rural retails surveys, Agriculture surveys etc. to make its database and this data is shared with various agencies, NGOS for research and policy making. Mr. Rajendra in his next slide explained about National Statistical Commission (NSC) which is an Apex organization and plays vital role in India. He notified that NSC was constituted in effect from 12th July 2006 with a mandate to evolve policies, priorities and standards in statistical matter. Speaker stated NSC functions as it decides short and long surveys, approves sample design, approves survey reports for release etc.

Speaker then briefed about **National Level Statistical Activities** in India. He stated few of the many activities out of which some are listed below:

Population Census carried by Registrar General and Census Commissioner,
Agricultural Statistics By DES
Education Statistics by HRD
Housing statistics by National Building Organization
Social Statistics and Environmental statistics both carried by NSO

Mr. Rajendra highlighted Job opportunities for students and participants in Statistical sector. He also indicated that this sector is huge with lots of job opportunities from officer level to junior level officers. He spoke about posts including ISS officers recruited by UPSC exams, SSS officers recruited by SSC exam, Contractual field Investigator. He also indicated jobs which are not just in government sector but also in private sector like Data Analyst, Data Scientist etc.

Mr. Rajendra concluded by indicating how Government of India is strengthening statistical system and creating awareness among people and students by conducting certain activities like presenting awards. He also informed that 2020 Statistics award is awarded to Mr. Rangarajan who was Chairman of NSC and RBI governor. He also told that various other activities like spot essay competition, organizing national seminars for researchers, planners, RBI etc. is carried out by N.S.O. Speaker in his last note informed about how government provides paid internship program for statistical students in NSO and encourages students to go to this sector as there are many job opportunities.

After the conclusion of the final session, a formal vote of thanks was given by Dr. P.V. Nimbalkar, Head of the Department of Computer Science and co-convenor of the webinar.

Around 380 delegates including academicians, professionals, individuals of different govt./semi-govt. organizations, researchers, students all over the nation and abroad attended this international webinar online and more than 5000 individuals from Assam, Jammu, Tamil Nadu, Kolkata, Dehradun, Varanasi, Chandigarh, Singapore etc viewed it on YouTube until now. The organizing committee of this webinar received valuable remarks like excellent webinar, excellent sessions, excellent speakers, relevant contents, well presented, nice coordination, gained worthy knowledge, gives a hope that nothing can stop us from learning, helpful for college students during lockdown period, good team work, excellent management, looking forward to such webinars in future etc.

The webinar was successfully carried out with the help of technical guidance and operations carried out by Mr. Nishit Jain from Tristar Software.