**Workshop Title:** Methods to Evaluate the Quality of Census and Sample Survey Data

**Organizers:**

Dr. Ankush Agrawal, Associate Professor, Indian Institute of Technology Delhi, New Delhi, India. Email: ankush@hss.iitd.ac.in

Dr. Vikas Kumar, Associate Professor, Azim Premji University, Bengaluru, India. Email: vikas.kumar@apu.edu.in

**Rationale:**

In the era of evidence-based policy, the evidence is crucial. While there have been attempts to generate and gather evidence, relatively little attention has been paid to the quality of evidence. This field of enquiry remains under-researched in most developing countries. In research methodology courses, which are a part of curricula in these countries, the discussion is largely confined to introduction to statistical techniques and tools and sources of data. There is hardly any discussion on the quality of data. Even training programmes and capacity building workshops do not pay attention to this issue. As a result, young researchers often do not consider it worthwhile to invest their time and efforts in examining data quality.

The proposed workshop seeks to sensitise researchers, young faculty members, and data journalists to the possible sources of inaccuracies and errors in censuses, which form the bedrock of the official statistical system, and sample surveys. The methods to evaluate the quality of data will form the mainstay of the discussion.

**Objectives and Goals:**

To familiarize the participants with the various data quality issues that arise with the census and sample surveys. We shall illustrate with the help of examples the errors that affect the census data (coverage and content errors) and sample surveys (sampling and non-sampling errors).

To introduce the participants to various techniques that can be used to evaluate the quality of data in census. We shall discuss internal consistency and external consistency checks to quantify the magnitude of coverage errors in census. The mechanisms to identify and quantify the content errors will also be discussed.

To acquaint the participants to the techniques to evaluate the errors in sample survey data. Our focus would be more on the non-sampling errors, which have been relatively under-researched. We shall also discuss how to quantify the impact of such errors on the survey estimates.
Expected outcomes and benefits:

Our public discourse is increasingly overwhelmed by data. The workshop will train the participants about tools and techniques to evaluate the data quality. It is expected that the attendees would be able to utilise these techniques to evaluate the quality of data collected in censuses and sample surveys of different countries. These techniques will be specifically relevant for evaluating the data quality in developing countries as the quality of data for several such countries is still deficient.

Target audience:

PhD students, young faculty members and researchers working in the areas of demography, economics, statistics, and sociology; data journalists

The number of participants:

We propose 25 to 40 participants. However, we are flexible regarding the intake.

Preferred days and duration:

We propose one-day workshop.

We do not have a preference regarding the day – we shall be available during the entire duration of the meeting. If the organizers (PAA) deem the workshop proposal fit to be floated, they can schedule it on any of the three days.

Format, Activities and Schedule:

We plan to have four sessions of one and a half hour each. In the first session, we shall introduce the participants to various data quality issues that affect the censuses and sample surveys. This will involve discussion on typologies and sources of errors.

The second and third sessions will discuss the methods to evaluate the census data quality. We shall talk about internal and external consistency checks, and how to use these to evaluate the quality.

In the last session, we shall discuss the non-sampling errors in the survey data and also the consequences of these errors on the survey results and the statistics of interest to policymakers.

The workshop will involve hands on training for the participants. We shall use the data from censuses and sample surveys conducted in India. The data would be shared before the workshop. It is best if the participants apply the techniques to the data parallelly.

First two sessions will be undertaken by Dr. Kumar and the last two by Dr. Agrawal. Both the instructors are familiar with all the topics being proposed in the workshop.
Whether or not you’d be willing to convert to an online workshop if necessary:

Given the nature of topics, it is best to hold the workshop in-person.

A/V needs beyond what is provided (see below):

None; participants may want to carry their laptop.

Bio-sketches:

Ankush Agrawal is an Associate Professor of Economics at the Indian Institute of Technology Delhi, India. His research interests include applied econometrics and development economics. Profile: https://hss.iitd.ac.in/faculty/ankush-agrawal

Vikas Kumar is an Associate Professor of Economics at the Azim Premji University, Bengaluru, India. His research interests include political economy and law and economics. Profile: https://azimpremjiuniversity.edu.in/people/vikas-kumar

Agrawal and Kumar have worked extensively on the quality of India’s official statistics. They are the authors of *Numbers in India’s Periphery: The Political Economy of Government Statistics*, Cambridge University Press (2020). The book examines the quality of census and survey statistics for India.