Topic: Tabulation of final data in NSSO, India

The unit level data in National Sample Survey (NSS) is finalised after running series of programmes at Data Processing Division (DPD). The important programmes are validation programmes(Phase-I/II/III), coverage programme, computer edit programme. After the unit level data is finalised, the multiplier values are calculated as per the sampling design. The multiplier values are posted in the unit level data. In each round of NSS survey a specified tabulation plan is prepared while finalising the objectives of the survey. The survey results in the form of tables are generated as per the tabulation plan. The steps of data tabulation can be summarised as below:

A. Calculation of multiplier:

In NSS survey design there is a concept of interpenetrating sub-samples. In every round, two independent samples are drawn as per the sampling strategy. In rural sector, Probability proportional size (PPSWR) is followed where size being total population of the village. In the urban sector, Simple random sampling without replacement (SRSWOR) is followed. The samples within a sub-sample are drawn independently and separate estimates can be obtained from each of the sub-samples. These sub-sample wise estimates are combined together to arrive at the final estimates. The final multiplier values are computed in a manner so that simple aggregation can generate the estimates. The precise steps may be summarised as below:

    i. Within each second stage stratum (SSS), the total no. of units available for survey, say 'H' and the no. of surveyed units, say 'h' are required from the unit level data. These values are obtained after running coverage programme. These 'H' and 'h' values along with the stratum size, which is separately available in estimation file, are used for calculation of multipliers.

    ii. The count of sub-sample wise (NSS) and sub-sample combined (NSC) values falling within a Stratum X SSS is obtained during the process of calculation of multipliers.

    iii. The different combinations of multipliers (MULT) as per the design are posted in the unit level data file along with the count of SSU's sub-sample wise and sub-sample combined.

iv. As per the aggregation method, the tabulation software takes care of the finer adjustments. Whenever a sample value is repeated in both the sub-samples, the multiplier value is (MULT/100), whereas if a sample value is present in only one sub-sample, the multiplier value is

v. (MULT/200). This is because the MULT value is expressed in 2 places of decimals.

B. Generation of workfile:

As an intermediate procedure for generation of tables as per tabulation plan, different types of work files are prepared. Work file, in other way, may be looked upon as an interface between the unit level data and the survey estimates. NSS data recorded in the form of different blocks of the schedule is digitised in the form of different levels having uniform text data structure. Work file is basically a uniform structure of data information culled out from different levels of the data file along with their identification parameters and multiplier components. Usually household/enterprise/item/person wise workfiles are prepared for obtaining different types of estimates. Hence understanding the concept of workfile preparation plays a pivotal role in generation of tables.

C. Generation of tables:

As per the requirement of the tables, workfile needs to be prepared. However, if a table requires the information which is available in only one level, then tables can be prepared from only that level of data. Basic requirement for table preparation is identification particulars (ID) including the break-up of tables (Sector, sex etc.) along with the data part. The in-house software developed in NSSO, DPD called 'DOSTAB' is ideally suited for this kind of tabulation.

D. Practical session on table generation:

The different components of the engine 'DOSTAB' needs to be learned through practical session only. Some of the tables from the tabulation plan to be prepared by each of the participants. Only then the entire table generation process will be fully understood.

****