Familiarization with the Unit Level Data
(Indian Experience)
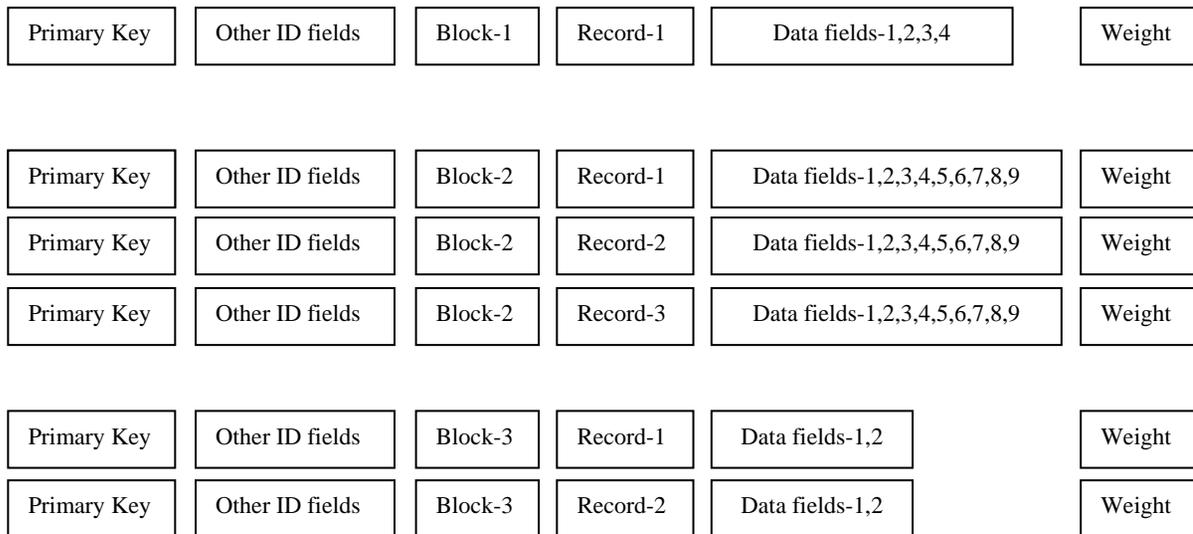
1. What is Unit Level Data ?

A large scale sample survey may be single-stage, two-stage or multi-stage. The selected samples in the first stage of sampling are called First Stage Units (FSU). In an areal survey, typically the FSUs are village in rural areas and urban-blocks in urban area. In the second stage, the selected samples are called Second Stage Units (SSU), which are normally households or enterprises. In the third stage, it may be persons within a selected household, and so on.

Unit Level Data refers to detail data in respect of the sampled units at the ultimate stage, along with the sampling weights for that stage. For example, in case of NSS household surveys, the lowest sampled units are households. Therefore, in this case unit level data means all data records in respect of each sample household, along with household-level multipliers posted in each record. It may be noted that household-level multipliers are composite weights of first and second stage sampling. So, on aggregation household-level multipliers will directly give the estimate for the whole country.

2. Structure of unit level data.

Data in respect of the ultimate stage units (e.g. households) may be collected in a number of blocks in the survey questionnaire. Data structure in each block are usually different. Hence, there will be separate data layout (or format) for different blocks. In database, data of different blocks are stored in different tables. But in text data, data of different blocks may be stored in separate block-level text files, or may be stored in a single large text file.

In order to retrieve all the data records for all the blocks in respect of a particular household, it is necessary that each table or block-level text file must have a common minimum set of identification fields, called primary key. So, the number of primary key fields must be same in all the blocks. Apart from these, there may be other secondary identification fields. In case of text data, it is advisable to keep the other ID fields common in all the records of all the blocks, so that processing becomes quick. But in a database, we may keep the other ID fields only in the Block-1.

Primary Key	Other ID fields	Block-1	Record-1	Data fields-1,2,3,4	Weight

Primary Key	Other ID fields	Block-2	Record-1	Data fields-1,2,3,4,5,6,7,8,9	Weight
Primary Key	Other ID fields	Block-2	Record-2	Data fields-1,2,3,4,5,6,7,8,9	Weight
Primary Key	Other ID fields	Block-2	Record-3	Data fields-1,2,3,4,5,6,7,8,9	Weight

Primary Key	Other ID fields	Block-3	Record-1	Data fields-1,2	Weight
Primary Key	Other ID fields	Block-3	Record-2	Data fields-1,2	Weight

In unit level data, Block number is called Level Number. Actually, each level represent a separate format of record. The number of data fields and records in different blocks (or levels) may be different. In case of text data

file, sometimes requisite amount of blank spaces are inserted in order to keep the record length fixed. The weight fields should be common in all the blocks.

3. Example.

To understand the unit level data of any survey, it is important to understand the sampling design and questionnaire of the survey. In NSSO surveys on household consumption expenditure, the first stage units (FSUs) are village or urban-blocks. Each sample FSU is given a unique 5-digited serial number. A large FSU is split into segments, and the selected segments are given serial number 1 or 2. Within each segment of a FSU, the households are classified based on certain characteristics (e.g. level of monthly consumption expenditure) and placed in groups which are called 2nd-stage stratum. These 2nd-stage strata are also given serial numbers 1,2,3 etc. Within each 2nd-stage stratum, households are sampled and selected households are given serial numbers 1,2,3 etc.

Therefore, in order to uniquely determine a sample household, the following identification fields are required :
(a) FSU serial number
(b) Segment number
(c) 2nd-stage stratum number, and
(d) Household serial number.

Hence, these identification fields constitute primary key for all the records of unit level data.

Apart from this, there are other secondary Identification Fields, which are totally dependent on the primary key. These may be, for example,

(a) Sector code (viz., 1 for Rural, 2 for Urban)
(b) State code

(c) District code
(d) Sub-division code
(e) Field Office code,

and so on. The description for these codes (i.e. names of State, names of District etc.) may be stored in a separate file outside the unit level data.

The data fields will depend on the block number. If a block contains information with respect to the household e.g. date of survey, respondent code, household size, type of household, total monthly consumption expenditure etc., it is enough to construct a single record for that block. In NSSO household consumption expenditure questionnaire, Block-1 and Block-2 have been merged into only one record, as they have very few data fields. Block-3 has been given a separate record.

But in Block-4, several information in respect of each member of the household are collected. Therefore, here it is necessary to construct one record for each person. For unique identification of a record, it is important to assign a record serial number (viz. 1,2,3,4….) to each record of Block-3. It may be noted that structure of each record within a block must be identical.

Similarly, in other Blocks, like 5.1, 5.2, 6, 7 etc. where quantity and value of many items of consumption are collected, it is necessary to construct multiple records.

****

Level-1
```
00011604062010111110101031111101101010000 211 160905191005 80                                    2   4      151608
```

Level-2
```
00111604062010111110101031111101101020000 50113561046106111        65 2    446772    01         2   4      151608
```

Level-3
```
00211604062010111110101031111101101030000111 48203 02         60                                 2   4      151608
00311604062010111110101031111101101030000222 42201 02         60                                 2   4      151608
00411604062010111110101031111101101030000351 18104 02         60                                 2   4      151608
00511604062010111110101031111101101030000451  9103 02         60                                 2   4      151608
00611604062010111110101031111101101030000552  5101 03         90                                 2   4      151608
```

Level-4
```
00711604062010111110101031111101101040101      35000    140001                                   2   4      151608
00811604062010111110101031111101101040102       7000     70001                                   2   4      151608
00911604062010111110101031111101101040108       3000     33001                                   2   4      151608
01011604062010111110101031111101101040117       2000     20002                                   2   4      151608
```