# Conducting Fieldwork and Survey Design

*(Prepared for Young South Asian Scholars Participating in SARNET Training Programme at IHD)*

Dr. G. C. Manna

# Structure of Presentation

- Advantages of Sample Surveys
- Some Important National Level Sample Surveys
- Designing a Survey: Some Important Points
- Orientation to Survey Methodology (NSS/ASI)
- Estimation Technique & Multipliers in the context of NSS & ASI
- Major Refinements in Sample Design of NSS
- Scope for Further Work on Sample Design

# Advantages of Sample Surveys

## Advantages over Complete Enumeration Survey / Census:

- Less time for fieldwork & release of results
- Cost effectiveness
- Better data quality by engaging much less personnel and imparting proper training
- Possible to provide population estimate within a tolerable (pre-specified) margin of error

# Some Important National Level Sample Surveys

- Household & Establishment Surveys by NSSO
- Annual Survey of Industries by CSO
- National Family Health Surveys, M/o HFW
- Recent Employment-Unemployment Surveys being conducted annually by Labour Bureau
- Census cum Surveys of Micro, Small & Medium Enterprises by DC (MSME)
- Income & Expenditure Surveys, India Human Development Survey (2004-05/'11-12) by NCAER

# Designing a Sample Survey: Some Important Points

- To have clear knowledge of target population
- Using appropriate sampling frame
  - ✓ Most relevant

    List frame or Area frame or Mix of both (examples)
  - ✓ Up-to-date (examples from ASI / NSS)
- Determining sample size to ensure desired level of precision (example)

# Designing a Sample Survey: Some Important Points …

- Adopting proper stratification
- Use of appropriate sampling technique for selection of units at different stages

  SRSWOR/SRSWR/PPSWR/CSS & Their Relative Efficiency
- No. of stages of sampling vis-à-vis cost & efficiency of the estimate
- Control of non-sampling errors

  Documentation of survey documents, Length of schedule, Choice of appropriate reference periods, Training to field personnel & data processing staff, Data validation & Publicity

# Orientation to
# Survey Methodology of NSS & ASI

# Scope & Salient Features of NSS

- Geographical coverage – whole country except a few inaccessible areas
- Continuing survey as successive '*rounds*'
- Methods of enquiry:
  - Interviewing respondents
  - Data as per books of A/Cs if available in case of Establishment Surveys

# Scope & Salient Features…

- Moving reference period (day/week/month/ year) – To reduce recall error and eliminate the effect of seasonality
- Survey duration mostly of one year / six months – Samples drawn in the form of sub-rounds; uniform spread of fieldwork within the sub-round
- Participation by States at least on an equal matching sample basis – Independent estimate as per ‘state sample’ as cross-check; Advantage of pooling
- Independent sub-samples – Easy computation of standard errors of estimates even in a complex design

# Scope & Salient Features…

- Sample size of recent rounds:

  12,000 – 16,000 PSUs; 10-20 Households/est. per PSU

  (Above sample size fairly adequate for many variables to provide state level estimates (but not true for rare events or those with higher variability)
- Permanent trained field investigators ensuring good quality data (now contract investigators also)

# Scope & Salient Features…

- Control of non-sampling errors

❑Two-tier training before fieldwork

❑Periodic meetings at the regional level

❑Training of data entry/processing staff

❑Documentation of various scrutiny programmes and their implementation

(Field offices and data processing centres of NSSO are in different locations across the country ensuring smooth execution of above)

# Some Important Subjects Covered (NSS 43rd Round:1987-88 onwards)

| **Subject** | **Round No.** |
|---|---|
| Consumer Exp. | Almost each round barring a few (large sample every 5 yrs; last: 68) |
| Empl.-unempl. | --- Do --- |
| Unorg. Manufacture | 45, 51, 56, 62, 67, 73 |
| Trading enterprises | 46, 53, 67, 73 |
| Service sector-Others | 57, 63, 67, 73 |

# Subjects Covered...

| **Subject** | **Round No.** |
|---|---|
| Survey on tribes | 44 |
| Housing / bldg. cons. | 44, 49, 58, 65, 69 |
| Education | 47, 52, 64, 71 |
| Disability | 47, 58 |
| LLH, Debt & Invest. | 48, 59, 70 |
| Slums | 49, 58, 65, 69 |

# Subjects Covered...

| **Subject** | **Round No.** |
|---|---|
| Health care | 52, 60, 71 |
| CPR | 54 |
| Travel habit | 43, 54, 65, 72 |
| Migration | 43, 49, 55, 64 |
| Aged persons | 42, 52, 58, 60 |
| Indian farmers | 59, 70 |

# Sample Design

- Broad sample design: Stratified two-stage (for smaller villages / urban blocks) or three-stage (larger villages/blocks):
  - ➢Primary Sampling Units (PSU):Villages/blocks
  - ➢Segmentation of large PSUs & selecting sample seg.
  - ➢Ultimate Sampling Units (USUs): Households/ establishments within each sample PSU

*Note that multi-stage design is the only available option because it is difficult to maintain up-to-date frame of all households/establishments*

# Sample Design…

- Sampling frame (PSU)

Household Surveys:

Rural: Census list of villages

Urban: Census list of EBs (till 15th round);

List of Urban Frame Survey blocks (thereafter)

Establishment Surveys:

Above with count of number of establishments/workers as per Economic Census to capture adequate number of establishments – particularly the bigger units

# Sample Design…

- Sampling frame (USU)

✓A list of USUs (i.e. households/establishments) prepared during survey by making a house to house visit within each selected PSU or selected segment(s) in case of large PSUs

✓Certain auxiliary information collected during the listing operation for stratification of USUs before sampling of USUs

# Sample Design…

- First-stage Strata (of PSUs)

Rural: District

Urban:

(1) Each M+ City – a separate stratum

(2)All other cities/towns within a district (from NSS 61st round) / NSS Region X Specified size class of towns (till 60th round and 65th round

# Sample Design…

- Sampling Procedure

PSUs:

Rural – Probability proportional to size (PPS)

Urban – Equal probability (PPS in Establishment Surveys with use of EC frame)

[Size measure in case of PPS: Population in HH Survey; No. of workers as per EC in Establishment Survey]

USUs:

Equal probability after appropriate stratification of households/establishments into a certain number of homogeneous groups i.e. strata

# A Few Examples of Stratification of HH

| Subject | Round | Stratification |
|---|---|---|
| Domestic Tourism | 65 | Rural: Structure type (Pucca / No pucca) x At least one member with overnight trip / Same day trip during last 30 days<br>Urban: MPCE class in place of structure |
| Housing Condition | 65 | Rural: Pucca/Semi-P/Others<br>Urban: MPCE Class |

# A Few Examples of Stratification of HH

| Subject | Round | Stratification |
|---|---|---|
| Survey of Farmers | 59 | 4 land possessed classes with the condition of some farming activity during last year |
| Land & Livestock Holding | 59 | Land possessed classes |
| Debt & Investment | 59 | Rural: Land holding classes x Indebted / Not indebted<br>Urban: MPCE classes x above |

# A Few Examples of Stratification of HH

| Subject | Round | Stratification |
|---|---|---|
| Disability | 58 | Mental disability (1); Speech/Hearing/Visual (2); Locomotor (3) |
| Morbidity, Health Care & Condition of the Aged | 60 | A member hospitalized during last 365 days (1); One child of age < 5 yrs (2); One member og age 60+ (3); Remaining HHs |

# Stratification of Establishments

**(See Reports on NSS Rounds 62, 63 & 67 for details)**

- Strata of own-account establishments by broad industrial category
- Strata of establishments with hired workers by broad industrial category
- Special efforts to form strata of larger units in terms of employment size or those with imprecise estimates of gross value added per worker

# Stratification in Annual Survey of Industries

- ASI covers registered manufacturing sector
- Units in frame grouped into 2 categories:

  Census sector (bigger units; W: 100+)

  Sample sector (other units)
- Sample sector units stratified by State/District X Industry groups
- Sample units selected from each stratum by circular systematic sampling after arranging them by number of workers

# Illustration on Sample Design: 68th Rd.

- Sampling frame: List of census villages/UFS blocks
- Stratification

  Rural: rural part of district

  Urban: Each million plus city; All other towns taken together within a district
- Sub-stratification: n/4 in number, each with equal population in case of rural and equal no. of HHs for urban (n: sample size for str.)

# Illustration from NSS 68th Rd…

- PSU Selection: PPSWR (R); SRSWOR (U)
- Segmentation of large PSUs and selection of 2 segments (one having max. population)
- Stratification of households:

Rural: Relatively affluent (1); Remaining with major earning from non-agr. (2), Others (3)

Urban: Top 10% of the households (1); Middle 60% (2); Bottom 30% (3)

(*Cut-off points in terms of average Monthly Per Capita Expenditure which varied across NSS Regions*)

# Estimation Procedure & Multipliers

- Use of design-based formula for estimation of totals/aggregates
- ‘Multipliers’ used to inflate sample values (which, in case of NSS multi-stage design, take into account the following: no. of PSUs surveyed; selection probability of PSU; no. of segments formed; no. of USUs listed & surveyed in the PSU/segment/ second stage stratum)
- Multipliers worked out at the level of PSU x segment x second stage strata in case of NSS
- Multipliers at the stratum level in case of ASI that adopts uni-stage design

# Estimation Procedure & Multipliers..

- Sample values multiplied by multipliers and then added up to derive first-stage stratum level estimates
- Stratum level estimates of aggregates added to obtain state (x sector) level estimates
- State totals give all-India estimate
- Ratios worked out at the last stage for the domain of interest

# Estimation Procedure & Multipliers..

$$\hat{Y}_j = \frac{Z}{n_{2j}} \sum_{i=1}^{n_{2j}} \frac{1}{z_i} \left[ D_i^* \times \frac{H_{i2j}}{h_{i2j}} \sum_{k=1}^{h_{i2j}} y_{i2jk} \right]$$

$$\hat{Y}_j = \frac{N}{n_{2j}} \sum_{i=1}^{n_{2j}} \left[ D_i^* \times \frac{H_{i2j}}{h_{i2j}} \sum_{k=1}^{h_{i2j}} y_{i2jk} \right]$$

# Estimation Procedure & Multipliers..

- Multipliers (M) in NSS Surveys (at PSU x segment x second-stage stratum level)

Rural Areas:

M = {(Stratum size) x (No. of segments formed in the PSU) x (No. of HHs listed in the PSU/ segment)} / {(No. of PSUs surveyed) x (Size of PSU) x (No. of segments surveyed in the PSU) X (No. of HHs surveyed in the PSU/ segment)}

Urban Areas:

M = (N*D*H) / (n*d*h)

- Multipliers in ASI (at stratum level)

M = (No. of units in the frame) / (No. of units surveyed)

# Major Refinements in Sample Design

- Overall sampling strategy broadly remained the same over NSS Rounds but certain refinements incorporated at different points of time

- Some major ones are described in subsequent slides

# Major Refinements…

- Sample size of PSUs increased manifold over time (from 1,000 plus in the initial rounds to around 12,000 – 14,000 now)
- Urban sampling frame of PSUs switched over from Census EBs till 15$^{th}$ round to NSSO's list of UFS blocks due to some shortcomings of EBs
- Economic Census initiated in 1977 to provide sampling frame for selecting PSUs in the Establishment Surveys to ensure adequate no. of units in the sample

# Major Refinements…

- Urban strata (of PSUs) changed to district since 61st round (2004-05) from NSS Region x Size class of towns earlier to facilitate generation of district level estimates for local planning

- Sampling method of PSUs changed to PPSWR (rural) / SRSWOR (urban) since 61st round from Circular Systematic Sampling (CSS) with PPS (rural) / Equal Probability (urban) earlier

# Major Refinements…

- Self-weighting design experimented in many rounds of 4–18 to save time & cost of tabulation
- Segment selection (large PSUs) in Establishment Surveys: Segment having max. concentration of units with prob. 1 and a sample from remaining at random introduced in Trade Survey (41$^{st}$ round: 1985-86) as compared to random sampling earlier

# Major Refinements…

- USU selection changed to SRSWOR from 61st round instead of CSS with Equal Probability earlier
- A combination of list frame of bigger units and usual area frame approach for other units experimented in 62nd and 63rd rounds of Establishment Surveys for improving the estimates of gross value added

# Scope for Further Work on Design

- Studying the relative efficiency of CSS with PPS/ Equal Probability adopted for a long time vis-à-vis PPSWR/SRSWOR as followed now
- Trying out modified sample design with due representation of:
  - Blocks with concentration of affluent people in surveys of HCE for improving estimates of MPCE
  - Large-sized households for improved estimate of number of persons
  - Bigger units/units with books of A/Cs for better estimate of GVAPW used in compilation of NAS

# Scope for Further Work…

- Exploring use of rotational panel design for repeat surveys of subjects like labor force, consumption exp. etc to give better estimates of change parameters
- Developing proper sample design for survey of rare events (drug abuse etc.)
- Pooling of central & state sample NSS data
- Using small area estimation techniques for estimates of smaller domains

Thank You