Kerala School of Mathematics
Course in Statistics for Scientists
Survey Sampling

T.Krishnan
Strand Life Sciences, Bangalore

6 March 2010

Types of Surveys

- Demographic surveys
- Consumption surveys
- Crop surveys
- Market surveys
- Opinion polls
- Election forecast surveys
- . . . .

Finite Population Sampling

- Random sampling—advantages
  1. Reduced cost
  2. Standard error (precision)
  3. Shorter time
  4. Prediction (Election prediction)
- Properties of a good survey
  1. Randomization
  2. Replication
  3. Use of auxiliary information
  4. Estimation of parameters
  5. Standard error of estimates

Principles of Sampling

- Randomization
  1. Selection of units with a probability scheme
  2. Removes bias
  3. helps calculate probability, confidence intervals, hypothesis tests
- Replication
  1. reduction of variance

Use of Auxiliary Information

- More control on units
- Improved schemes
- Scheme depends upon type of auxiliary information
- Suitable estimates for schemes
- May be used in sampling and/or estimation

Standard Sampling Schemes

- Simple Random Sampling
  1. with replacement
  2. without replacement
- Systematic Sampling
  1. linear
  2. circular
  3. problem of variance estimation
- Stratified Sampling
  1. sampling schemes in strata
  2. allocation of sample size to strata
- Multistage Sampling
- Cluster Sampling
- Unequal Probability Sampling
  1. PPS Sampling

Estimation Methods

- Unbiased estimation
- Variance estimation
- Ratio estimation
- Regression estimation
- Bayesian estimation

Simple Random Sampling With Replacement (SRSWR)

- Population size: N
- Sample size: n
- Number of possible samples: $\binom{N}{n}$
- Equally likely
- Method: One by one with replacement
- Use of Random Numbers
- Population Values: $Y_1, Y_2, \ldots, Y_N$ with total Y and mean $\bar{Y}$
- Population Variance σ^2
- Sample Values: $y_1, y_2, \ldots, y_n$ with sample mean $\bar{y}$ and sample variance s^2

Estimation in SRSWR

- $\hat{\bar{Y}} = \bar{y}$
- $V(\hat{\bar{Y}}) = \frac{\sigma^2}{n}$
- $\hat{V}(\hat{\bar{Y}}) = \frac{\hat{\sigma}^2}{n}$
- Approximate 95% confidence for $\bar{Y}$:

$$\bar{y} \pm 1.96 \frac{\hat{\sigma}}{\sqrt{n}}$$

Sample Size Determination

- If Y is a (0, 1) variable
- $\bar{Y}$: Population proportion P
- $\bar{y}$: Sample proportion p
- $\sigma^2 = P(1-P)$, largest for $P = \frac{1}{2}$
- Closer P is to $\frac{1}{2}$ the larger the required n
- Estimate to be within 0.02 of the actual value of P with probability 0.95
- Worst case: $P = \frac{1}{2}$
- $\frac{1.96}{\sqrt{4n}} = 0.02$ gives $n = 2500$.