

Bayesian Inference for Surveys

Module 7 (Continued)

Models for Stratified Sample Design

Model Refinements

- A popular design: 2 units sampled per stratum
- Too few to estimate the variance
- Option 1
 - Pooled variance

$$Y_{ih} \mid \mu_h, \sigma \sim iid N(\mu_h, \sigma^2)$$

$$\pi(\mu_1, \dots, \mu_H, \sigma) \propto \sigma^{-2}$$

$$i = 1, 2, \dots, N_h; h = 1, 2, \dots, H$$

Implementation

- Missing Data Approach
 - Create Two Variables Y, Z (Strata indicators)
 - Set all the unobserved values to missing
 - Multiply Impute the unobserved values using a regression model of Y on dummy variables based on Z
 - This is akin to fitting a one-way analysis of variance with Strata as Groups

- Option 2
 - Proper prior for variance parameters across strata

$$Y_{ih} \mid \mu_h, \sigma_h^2 \sim iid N(\mu_h, \sigma_h^2)$$

$$\pi(\mu_h, h = 1, 2, \dots, H) \propto 1$$

$$\sigma_h^{-2} \sim iid Gamma(a, b)$$

$$a, b : Known$$

- Option 3
 - Random effects on both Mean and Variance

$$Y_{ih} \mid \mu_h, \sigma_h^2 \sim iid N(\mu_h, \sigma_h^2)$$

$$\mu_h \mid \sigma_h^2 \sim N(\mu, c_h \sigma_h^2)$$

$$\sigma_h^{-2} \sim iid Gamma(a, b)$$

$$c_h, a, b : Known$$

Options 2 and 3 require Gibbs sampling approach and can be implemented using Openbugs, Stan or Winbugs, Proc MCMC etc

Systematic Sampling

- Population size $N=nk$
- Sample size= n
- Elements sequenced into n groups each of size k
- Choose a random number between 1 and k (say, L)
- Sample : $L, L+k, L+2k, \dots, L+(n-1)k$

Can be viewed as sampling 1 element from each of the n strata of size k . However, only one random start determines all the selection

(Refinements for N that is not multiple of n are available)

Model

- Treat as SRS
- Combine adjacent 2 groups to create $n/2$ strata with 2 selections per stratum
- Assume some model for Y as a function index (ordered values)

(Variance estimation is also a problem in the design based inference)