EXPOSITORY PAPER

ARASH FARAHMAND

Abstract.

Contents

1. Introduction 2
2. Sample Design and Survey Data 2
2.1. Types of Sampling 2
2.2. Sample Design 3
3. Complexity of Analyzing Survey Data 3
4. Strategies for Variance Estimation 3
4.1. Replication Sampling: A General Approach 3
4.2. Balanced Repeated Replication (BRR) 3
4.3. Jackknife Repeated Replication (JRR) 3
4.4. The Bootstrap Method 3
4.5. The Taylor Series Method (Linearization) 3
5. Preparing for Survey Analysis 3
6. Conducting Survey Data Analysis 3
7. Conclusion 3
References 4

Date: March 18, 2008.

1. Introduction

2. Sample Design and Survey Data

In this paper we only deal with *probability sampling*, in which each individual has a known (nonzero) probability of being included in the sample. One way of achieving this goal is to implement a multistage approach in sampling.

2.1. **Types of Sampling.**

2.1.1. *Simple Random Sampling (SRS).* This is the simplest sample design. Each individual has an equal chance of being selected in the sample from a known list of all individuals in the population. Two major types are with and without replacement. The other sampling methods are modifications of SRS.

2.1.2. *Systematic Sampling.* This method selects every k^{th} individual into the sample after a random start. Different sets of elements may have different inclusion probabilities, thus complicating the calculation of variance estimate.

2.1.3. *Repeated Systematic Sampling.* To overcome the problems in systematic sampling, we select several smaller systematic samples, going down the list several times, with a new starting point in each pass. We can easily estimate the variance from the variability of separate estimates from each sample.

2.1.4. *Stratified Random Sampling.* In this type we first divide the population into several strata and we take samples within each stratum separately. There are several benefits of using this type of sampling, among them reducing the sampling variance, getting separate estimates for each stratum, and better organization of the administrative work within each stratum. The first step in stratified random sampling is to calculate the statistics for each stratum. The second step is to combine the statistics of strata. Sampling within each stratum is usually SRS.

2.1.5. *Cluster Sampling.* We use cluster sampling when we sample by groups or clusters rather than by individuals. This type has different forms: one-stage, two-stage, or multistage. In one-stage cluster sampling, we sample clusters with probability proportional to the size of the clusters (PPS sampling), or probability proportional to the estimated size (PPES sampling). In two-stage cluster sampling, we first divide the population into smaller units, then we take a sample of these units before taking a sample within each of the chosen units. This technique is also called subsampling.

2.1.6. *Multistage Sample Design.* In multistage cluster sampling, the sampling units are groups of elements except for the last stage. When we use stratified sampling and unequal-sized clusters, the estimation process gets complicated.

2.1.7. *Unequally Weighted Sampling.* Sampling with equal probabilities (weights) has three major shortcomings:

(1) The total number of sample units would be proportional to the number of individuals in the sample, so estimators may have large variance.
(2) Administering this kind of sample may be difficult.
(3) The cost of sample is not known in advance.

Using unequally weighted sampling relieves the sampling from the above three problems. The sampling variance decreases by assigning unequal probabilities to sampling units in different strata.

2.2. **Sample Design.**

3. Complexity of Analyzing Survey Data

4. Strategies for Variance Estimation

4.1. **Replication Sampling: A General Approach.**

4.2. **Balanced Repeated Replication (BRR).**

4.3. **Jackknife Repeated Replication (JRR).**

4.4. **The Bootstrap Method.**

4.5. **The Taylor Series Method (Linearization).**

5. Preparing for Survey Analysis

6. Conducting Survey Data Analysis

7. Conclusion

References

[1] Chen, Sean X., *Weighted Polynomial Models and Weighted Sampling Schemes for Finite Population*, The Annals of Statistics, Vol. 26, No. 5., 1894-1915, 1998.

[2] Cochran, William G., *Sampling Techniques*, John Wiley and Sons, 1977.

[3] Graubard, Barry I. and Fears, Thomas R., *Standard Errors for Attributable Risk for Simple and Complex Sample Designs*, Biometrics 61, 847855, 2005.

[4] Kalton, Graham, *Introduction To Survey Sampling*, Sage Publications, 1983.

[5] Lee, Eun Sul and Forthofer, Ronald N., *Analyzing Complex Survey Data, Second Edition*, Sage Publications, 2006.

[6] Lo, Albert Y., *A Bayesian Method for Weighted Sampling*, The Annals of Statistics, Vol. 21, No. 4., 2138-2148, 1993.

[7] Lohr, Sharon L., *Sampling: Design and Analysis*, Brooks/Cole Publishing Company, 1999.

[8] The National Health and Nutrition Examination Survey (NHANES), *Analytic and Reporting Guidelines*, http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/analytical_guidelines.htm, 2006.

[9] Research Triangle Institute, *SUDAAN User's Manual, Release 8.0*, Research Triangle Institute, 2001.