UNIVERSITY OF CALIFORNIA

Los Angeles

# Some Estimators of Subuniverse Means for Use With Lattice Sampling

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biostatistics

by

David Elvin Abbey

Doctoral Committee:

Professor Raymond J. Jessen, Chairman

Professor Abdelmonem A. Afifi

Professor Virginia A. Clark

Professor Max. R. Mickey, Jr.

Professor N. Donald Ylvisaker

1972

The dissertation of David Elvin Abbey is approved, and it is acceptable in quality and form for publication on microfilm.

Committee Chairman

University of California, Los Angeles

1972

# TABLE OF CONTENTS

Page

Page

## Acknowledgements

I wish to express my deepest appreciation and gratitude to Professor R. J. Jessen for guiding me in this research with invaluable suggestions, constructive criticisms and many helpful discussions.

I am also grateful for the help and advice given so freely by the other members of the committee--Professor A. Afifi, Professor V. Clark, Professor M. Mickey, and Professor N. Ylvisaker. I wish to especially thank Professor A. Afifi for his help with the matrix algebra for the least squares estimator, and Professor R. Jennrich who gave me invaluable assistance with the proof of the unbiasedness of the least squares estimator.

Computing assistance was obtained from the Health Sciences Computing Facility, UCLA, sponsored by NIH Special Research Resources Grant RR-3.

Finally, I wish to thank Linda Bundick and my wife, Judy for typing this thesis.

"Giving thanks always for all things unto God." Ephesians 5:20.

VITA

November 19, 1943 -- Born, Sidney, B.C., Canada

1967 -- B.Sc., Walla Walla College, Walla Walla, Washington

1967 - 1968 Research Assistant, Department of Physiology and Biophysics, Loma Linda University, Loma Linda, California

1968 - 1969 Research Assistant, Department of Biostatistics, School of Public Health, Loma Linda University Loma Linda, California

1969 - 1972 National Institute of Health Trainee, Division of Biostatistics University of California, Los Angeles

ABSTRACT OF THE DISSERTATION

Some Estimators of Subuniverse Means

for Use With Lattice Sampling

by

David Elvin Abbey

Doctor of Philosophy in Biostatistics

University of California, Los Angeles, 1972

Professor Raymond J. Jessen, Chairman

In many sample surveys in addition to having estimates for the entire universe it is desirable to have intra-universe (or subuniverse) estimates where the sample sizes within many of the subuniverses may be too small to give adequate accuracy with direct estimators. It may be possible to increase the accuracy of subuniverse estimates by making use of the structure of the population to "borrow" information from sample units outside of the subuniverse.

As a means of doing this three alternative linear models or classifications of the universe--0-way, 1-way, and 2-way, and the use of a form of 2-way stratified sampling known as lattice sampling were considered. In the two-way stratification considered, the subuniverses are the "rows", the ordinary strata are the "columns" and the "cells" (intersections of rows and columns) of the two-way classification are the primary sampling units. In lattice sampling only some of the cells are selected subject to the restriction that fixed numbers of cells be

sampled in each row and column.

Estimators for intra-universe means were developed under each of the three models using a number of different estimation methods. For the case when the cells in the two-way stratification are of equal size and are selected with equal probability according to a two-stage lattice sampling scheme, the sampling variances and biases of the estimators were derived. The schemes used were the "simple" (which used simple sample averages to estimate the parameters), the least squares, the parametric (which used the parameter estimates for every cell), and the missing-cell scheme (which used the observed sample means in sampled cells and parameter estimates in missing cells). Those schemes which led to estimators having the lowest mean square error averaged over the sub-universes were termed best. Since these mean square errors depend largely on the structure of the populations dealt with, empirical tests were carried out on several small, synthetic populations.

It was found that:

1) The estimators using the 2-way model are never worse and are usually better than those using the 1-way model.
2) The 2-way model is better than the 0-way model unless sub-universe differences are very small or the within cell variance is large compared to the between cell variance.
3) The simple method is better or at least as good as the least squares method except when the 2-way model fits the data very closely.
4) The missing cell methods were better than the parametric methods when used with the 2-way model, the same for the

1-way model, and usually worse for the 0-way model.

5) The best combination of methods to use is usually the 2-way model with the simple missing cell techniques.

Estimation schemes based on lattice sampling were compared with those based on two alternative one-way stratification schemes. It was found analytically that the lattice methods could never have a larger mean square error than the other methods and that substantial gains in accuracy could occur.

The form of the estimators was considered for the more general case where cells are of unequal size and are selected with unequal probability. The matrix formulation for the least squares estimators was extended to this case and it was found that when cells are selected with probability proportional to size the least squares estimates became self-weighting.

When auxiliary variables are available the estimators can be easily extended so that multiple regression techniques can be used to exploit this information also. It was indicated how this could be done with the 2-way least squares estimator and one auxiliary variable.