# The Effect of Clustering on the Precision of Estimation

A thesis presented in partial
fulfilment of the requirements

for the degree

of Master of Business Studies
in Marketing at
Massey University

Zhengping Guan

1997

**ABSTRACT**

The effect of clustering interval on design effect may be important in selection of alternative sampling designs by evaluating the cost-efficiency in the context of face-to-face interview surveys. There has been little work in investigating this effect in New Zealand. This study attempts to investigate this effect by using data from a two-stage sampling face-to-face interview survey. Seventeen stimulated samples are generated. A simple method, $design \;\; effect = \frac{ms_b}{ms}$, is developed to estimate design effects for 81 variables for both the simulated samples and the original sample. These estimated design effects are used to investigate the effect of clustering interval. This study also investigates the effect of cluster size. The results indicate that clustering interval has little influence on design effect but cluster size substantial influence. The evaluation of the cost-efficiency in alternative clustering intervals is discussed. As an improvement in the efficiency of a sample design by an increase in clustering interval can not be justified by the increase in cost, it seems that the sample design with the smallest clustering interval is the best. An alternative method $design \;\; effect \approx mr^2$ is also discussed and tested in estimating design effects. The result indicates that the applicability of $design \;\; effect \approx mr^2$ is the same as that of $design \;\; effect = \frac{ms_b}{ms}$.

**ACKNOWLEDGMENT**

I would like to thank ACNielsen McNair for providing data from a face-to-face interview survey.

Thanks are also due to Mr Nick Jones, Managing Director of ACNielsen McNair, for his helpfulness and kind cooperation, and due to Mr James Reilly for his kind assistance in preparing the data.

**CONTENTS**

Page

## LIST OF TABLES

Page

## LIST OF FIGURES

Page

## 1. INTRODUCTION

Surveys using clustered multi-stage sampling designs are common in research in business and other social sciences. For a given sample size, these sampling designs may reduce the cost of data collection. However, such designs lead to increase in the sampling variances of estimates.

This study investigates the way in which final stage clustering affects sampling variances in face-to-face interview surveys.

In view of the need to make an adjustment to a sampling variance estimate from a complex sample design, Kish (1965) proposed a measurement which he called "design effect" to describe the sampling variance increase due to the complex sample design. He held the position that sample designs affect variance estimation and statistical analysis. However, Skinner, Holt & Smith (1989 chapter 2) argued that it was population structure rather than sample designs that affected variance estimation and statistical analysis. These two positions are often consistent. For a given sample design, population structure may affect variance estimation and statistical analysis, and vice versa.

Skinner et al (1989, p 24) also proposed an alternative measurement which they called "misspecification effect" instead of design effect. That is, the measurement of sample design efficiency is sampling variance of the actual sample design over the expected value of sampling variance of a simple random sample with the same size, rather than sampling variance of the actual sample design over sampling variance of a simple random sample with the same size. However, it is difficult in practice to obtain the expected value of a sampling variance estimate. Thus, design effect is likely to be more applicable in measuring the efficiency of sample designs than misspecification effect.

Sampling variance increase due to clustering in surveys is caused by similarity of elements within clusters. This similarity is measured by the homogeneity of within-cluster elements.

There is a voluminous body of literature concerning complex sample design, variance estimation, design effect and homogeneity. However, there has been little research into the relation between design effect and intervals of selecting elements within clusters in New Zealand. The need to evaluate the cost-efficiency of the alternative sample designs with different clustering intervals requires to conduct an investigation into the effect of clustering interval on design effect.

Data for this study is from a face-to-face interview survey conducted by ACNielsen-McNair. This is a two-stage sample (see Chapter 4 for specification of the sample). A number of simulated samples are drawn from it to investigate the effect of clustering interval (see Chapter 4 for the detailed discussion in generating simulated samples).

Based on the design effects estimated from both the original sample and the simulated samples, this study investigates the following:

a. *The relation between design effect and clustering interval;*

b. *The relation between design effect and cluster size;*

c. *The applicability of the formula:*

$$design \quad effect \approx mr^2$$

*(see Chapter 4 for both specification and derivation of this formula);*

d. *The effect of clustering interval on cost-efficiency of alternative sample designs.*

The results for both *a* and *b* should be that design effect decreases with either increase in clustering interval or decrease in cluster size. The result for *c* should justify the alternative estimation method for design effect. The result for *d* should provide the guideline for selection of the alternative sample designs with different clustering intervals.