

Imputation Methods for the Population and Housing Census 2000 in Korea

Ryu, Jea-Bok
*Professor, Dept. of Statistics,
Chongju University
Chongju, Korea.*
jbryu@chongju.ac.kr

Kim, Young-Won
*Professor, Dept. of Statistics,
Sookmyung Women's University
Seoul, Korea.*
ywkim@sookmyung.ac.kr

Park, Jin-Woo
*Assistant Professor, Dept. of Applied Statistics,
The University of Suwon
Whasung, Korea.*
jwpark@mail.suwon.ac.kr

Lee, Jae-Won
*Statistician, Population Census Division,
National Statistical Office
Daejeon, Korea*
jwlee@nso.go.kr

1. Introduction

The Korea National Statistical Office (NSO) conducted a pilot survey of the Population and Housing Census 2000 in November 1999. The main purpose of the pilot survey was to pretest questionnaires and survey procedures to identify problems prior to the Census. The pilot survey was executed to get 22 items for the complete enumeration survey and 34 additional items for the sample survey. For the complete enumeration survey, 1058 enumeration districts(EDs) from 16 different provinces were selected and 105 EDs were used for sample survey to pretest the whole Census procedure.

As a result, 8 items showed relatively high non-response rate among the 56 items. Especially, the nonresponse rates of the item on the total area of floor space in housing part and the item on marital status in population part were very high. The Korea NSO is planning to do three times callbacks to fill out nonresponses. The Korea NSO now want to check whether it is possible to replace the last two call-back procedures with the imputation method suggested by this study while retaining the first call-back procedure which include predetermined consistency checks.

In this study, we will suggest imputation strategies for the items which may have relatively high nonresponse rates in Census. Hence we concentrate solely on the missing total area of floor space and marital status. Using pilot survey data for census 2000, we will suggest the efficient classification method to build up the imputation cells. For this purpose, Classification and Regression Tree(CART) method proposed by Breiman, et al.(1984) was employed. Several imputation methods are then discussed.

2. Imputation Procedure

In the pilot survey, the nonresponse rates of total area of floor space and marital status were about 2.9% and 2.6% respectively after the first callback. First of all, a construction method for appropriate imputation cell was considered using CART method. For total area of floor space, various covariates such as region, number of rooms, total area of site and number of resident households are used to generate imputation cells. The optimal number of imputation classes turned out to be nine. Also we build up 6 imputation cells for marital status with covariates such as age, sex and region.

In this study, a hot-deck procedure which imputes a value using data from the nearest neighborhood within the imputation cell and the model based approach are applied to impute the missing values and then performance of imputation methods are compared.

3. Evaluation of Proposed Imputation Method

For the Census 2000 in Korea, three times callbacks are originally planned to fill out non-responses. Consequently four editions of survey data sets are created in pilot survey, one from the initial survey and the other three modified data sets from each callback procedure, respectively. The Korea NSO wants to study whether the last two callbacks can be replaced with imputation.

In order to evaluate the methods proposed in this article, we applied imputation methods for the missing values in the data set after the first callback and then compared the imputed values with the associated observed values appearing in the final data set after the third callback. Effects of the imputation methods on the point estimation are negligible because of a minor degree of nonresponse after the first callback. The results suggest that the last two callbacks can be omitted and imputation after only one callback is strongly recommendable.

REFEFENCE

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984), "Classification and Regression Trees", Chapman & Hall.

Thibaudeau, Y., Williams, T. and Krenzke, T. (1997), "Multivariate Item Imputation for the 2000 Census Short Form", in Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 371-376.

Williams, T. R. (1998), "Imputing Person Age for the 2000 Census Short Form: A Model-Based Approach", Technical Report, Bureau of the Census Statistical Research Division.