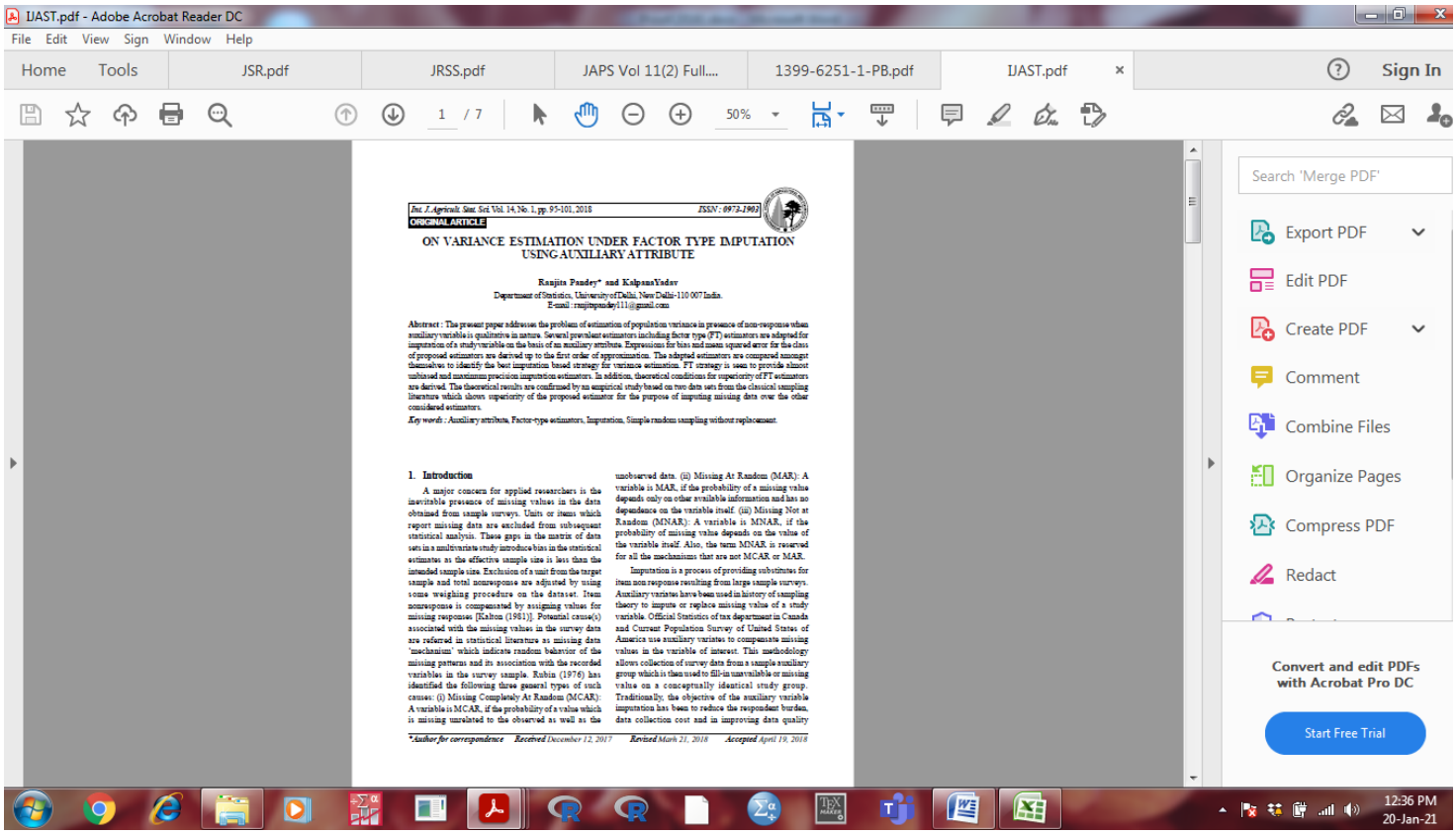

Int. J. Agricult. Stat. Sci. Vol. 14, No. 1, pp. 95-101, 2018 ISSN : 0973-1903

ORIGINAL ARTICLE

ON VARIANCE ESTIMATION UNDER FACTOR TYPE IMPUTATION USING AUXILIARY ATTRIBUTE

Ranjita Pandey* and KalpanaYadav
Department of Statistics, University of Delhi, New Delhi-110 007 India.
E-mail : ranjitapandey111@gmail.com

Abstract : The present paper addresses the problem of estimation of population variance in presence of non-response when auxiliary variable is qualitative in nature. Several prevalent estimators including factor type (FT) estimators are adapted for imputation of a study variable on the basis of an auxiliary attribute. Expressions for bias and mean squared error for the class of proposed estimators are derived up to the first order of approximation. The adapted estimators are compared amongst themselves to identify the best imputation based strategy for variance estimation. FT strategy is seen to provide almost unbiased and maximum precision imputation estimators. In addition, theoretical conditions for superiority of FT estimators are derived. The theoretical results are confirmed by an empirical study based on two data sets from the classical sampling literature which shows superiority of the proposed estimator for the purpose of imputing missing data over the other considered estimators.

Key words : Auxiliary attribute, Factor-type estimators, Imputation, Simple random sampling without replacement.

1. Introduction

A major concern for applied researchers is the inevitable presence of missing values in the data obtained from sample surveys. Units or items which report missing data are excluded from subsequent statistical analysis. These gaps in the matrix of data sets in a multivariate study introduce bias in the statistical estimates as the effective sample size is less than the intended sample size. Exclusion of a unit from the target sample and total nonresponse are adjusted by using some weighing procedure on the dataset. Item nonresponse is compensated by assigning values for missing responses [Rubin (1981)]. Potential cause(s) associated with the missing values in the survey data are referred in statistical literature as missing data 'mechanism' which indicate random behavior of the missing patterns and its association with the recorded variables in the survey sample. Rubin (1976) has identified the following three general types of such causes: (i) Missing Completely At Random (MCAR): A variable is MCAR, if the probability of a value which is missing unrelated to the observed as well as the unobserved data. (ii) Missing At Random (MAR): A variable is MAR, if the probability of a missing value depends only on other available information and has no dependence on the variable itself. (iii) Missing Not at Random (MNAR): A variable is MNAR, if the probability of missing value depends on the value of the variable itself. Also, the term MNAR is reserved for all the mechanisms that are not MCAR or MAR.

Imputation is a process of providing substitutes for item non response resulting from large sample surveys. Auxiliary variates have been used in history of sampling theory to impute or replace missing value of a study variable. Official Statistics of tax departmentin Canada and Current Population Survey of United States of America use auxiliary variates to compensate missing values in the variable of interest. This methodology allows collection of survey data from a sample auxiliary group which is then used to fill-in unavailable or missing value on a conceptually identical study group. Traditionally, the objective of the auxiliary variable imputation has been to reduce the respondent burden, data collection cost and in improving data quality

**Author for correspondence Received December 12, 2017 Revised March 21, 2018 Accepted April 19, 2018*