

## TREATMENT OF MISSING DATA VIA MULTIPLE IMPUTATION AND FINITE-DIFFERENCE INTERPOLATIONS

G.A. DAWODU

---

Department of Statistics, College of Natural Sciences, University of  
Agriculture, Abeokuta, Ogun State, Nigeria  
E-mail: agad1963@yahoo.com

---

### ABSTRACT

The missing data problem has been under study for sometime now, this work explains the state of the art and introduces another data mining technique to further handle the missing data problem. Numerical recipes are not new to the study of the problem, this work is an exhibition of more numerical recipes and the types of data they are appropriate for.

**Keywords:** Data mining, multiple imputation, principal components, covariates and numerical recipes.

### INTRODUCTION

The sources of missing data are numerous, infact they constitute an inexhaustive list (Dawodu, 2004). Until recent times, the ways through the problem is usually handled have been found to affect the survey adversely (Friedman, 1997). In the past, the data Analyst merely disregards all questionnaires that were not returned or whose questions were not all responded to or carry-out a posterior survey. These two ways have been found to, either affect the precision of the parameters of interest or, be uneconomical and time consuming (Dawodu, 2004). Multiple imputation has been identified, as an area in data mining that can help profer solutions to those problems. We shall view it as the combi

nation of two or more single imputations (User's manual, 2001). Now data exists in various forms and patterns and so also are the "holes" to be filled in them. Two major categories are:

1. Two rows or columns data, where one row (or column) is filled (X), the second row (or column) contains holes (Y). X is usually the, principal component, covariate, independent variables (e.g. a set of "timing measures") e.t.c.
2. Multiple rows or columns data, where at least, one row (or column) is filled (covariates) and others having holes.