THREE MODIFICATIONS TO THE AARDVARK STATISTICAL PACKAGE IMPLEMENTED AT KANSAS STATE UNIVERSITY

by

KRISTOPHER LEE ARHEART

B. S., Kansas State University, 1970

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:

Tom L. Gallagher
Major Professor

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION

AARDVARK[1] is a modular statistical package developed at Iowa State University to provide analysis of variance and covariance for a broad scope of statistical problems. In particular, AARDVARK distinguishes two basic types of statistical models.

1. Analysis of variance and covariance for models of the balanced complete structures class.
2. General multiple linear regression analysis models.

Balanced complete structure models are those having an equal number of observations for every combination of levels for each factor. AARDVARK provides an analysis of variance for cases in which some factors are nested in the levels of other factors or in combinations of other factors. Therefore, AARDVARK can analyze factors which are completely crossed, completely nested, or a combination of crossed and nested. Analysis of covariance adds the capability of adjusting each factor for covariates either jointly or separately at the user's discretion. An analysis of variance on the covariates is also offered.

AARDVARK is designed with user convenience as a goal. It provides, in one package, a set of routines that performs analysis of variance and covariance on various statistical models. AARDVARK allows problems to be formulated in terms that are consistent with the natural statistical representation of the problem. The user is offered options to control

---

[1] AARDVARK is an acronym standing for analysis of variance system, algebraic model options, residual and means options, data format options, variate and covariate analysis, analysis on means option, requested pooled terms and key statistical transformations.

the printing of certain statistics, the type of input data, and certain model conditions. A set of standard mathematical data transformations is also provided as well as the ability to include user prescribed transformations.

Since its implementation at Kansas State University in 1968, AARDVARK has undergone changes to correct errors and to provide better service for its many users. This report will discuss three recent modifications. First, the correction of the routine that provides F statistics for the analysis of variance of mixed or random effects models will be discussed. The next topic will be the implementation of a set of routines that selects the best subset of predictor variables for a linear regression analysis. Providing the user with clear, intelligible error diagnostics instead of the usual error numbers will be discussed last.

CHAPTER 2

CORRECTION OF THE ERRONEOUS F TEST DENOMINATORS AND COMPONENTS OF VARIANCE FOR TERMS IN A MIXED OR RANDOM EFFECTS MODEL ANALYSIS OF VARIANCE

The major objectives of an analysis of variance for a random effects model are to estimate the variance components and to test hypotheses about their magnitude by using an F test. To compute estimates of variance components, an expected mean square for the denominator of the F statistic must be obtained. This is also true for testing hypotheses about variance components. Therefore, the denominator of the F statistic is of prime importance in an analysis of variance with random effects. In AARDVARK, random and mixed models are handled similarly, so only random effects models will be discussed.

In a random effects analysis of variance model with two factors, there are four mean squares: the A, the B, the AB interaction, and the error. Figure 1 illustrates a two factor random effects analysis of variance table. The expected mean square of the error is composed of only the error variance. The expected mean square for the AB interaction contains the error variance and a coefficient times the AB variance component. The coefficients for the AB interaction variance component, the A factor variance component, and the B factor variance component are the respective number of observations on each effect. The A and the B expected mean squares are composed of the error variance, the AB interaction variance times its coefficient, n, and the A or the B variance times their respective coefficients, bn, and an. For the coefficients above, a is the number of levels of factor A, b is the number of