THE EFFECTS AND DETECTION OF COLLINEARITY IN AN ANALYSIS OF COVARIANCE

by

JO JANE GIACOMINI

B.S., Colorado State University, 1977

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1982

Approved by

James J. Higgins
Major Professor

ACKNOWLEDGMENTS

I wish to thank the members of my committee, Dr. James Higgins, Dr. George Milliken and Dr. Michael Deaton. Special appreciation is extended to Dr. James Schwenke for his valuable suggestions during the course of this research.

I also wish to thank the entire Statistics Department for making my stay at Kansas State a pleasant experience. I am also thankful to Sharon Luce for doing an excellent job of typing this report.

Finally I wish to thank my family and friends for their encouragement and understanding.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Analysis of covariance is used to reduce the variability due to experimental error and thus increase precision in randomized experiments. The procedure is to measure a covariate, a source of variation that cannot or has not been controlled by the experimental design, along with the response variable. This covariate, X, is related to the response variable, Y. The relationship between X and Y may be different for each level of the treatment giving different regression lines. Through the analysis of covariance, a comparison of these regression lines is made.

The purpose of the analysis of covariance is to test whether or not there are significant distances between the regression lines, or in more common terminology, whether or not there are treatment effects. The difference between analysis of covariance and the analysis of variance is that analysis of variance is based on the unconditional distribution of Y, the response variable, whereas analysis of covariance is based on the conditional distribution of Y given X, the covariate. The analysis of covariance model combines into one model, both the regression and analysis of variance models, by having different regression models for each population.

In the analysis of covariance, different testing strategies may be employed to determine an appropriate form of the covariance model. One strategy may assume the covariate is necessary and that the slope parameter is constant across all treatments. The only test of interest is one for equal intercepts of the parallel lines. Another strategy might initially test whether or not the covariate is necessary, then based on the conclusion of that test do another test.

One of the assumptions made in analysis of covariance is independence of covariate and treatment, or noncollinearity between covariate and treatment. When a treatment is adjusted for the covariate, it is as if the treatment mean is changed to the value it would be expected to have if all observations had the same covariate value. However, if the treatment affects the covariate, the adjustment removes part of the treatment effect and may lead to incorrect conclusions.

If the design of the experiment is simple, that is, a one-way analysis of covariance with one covariate, then the problem of a collinearity between treatment and covariate can be detected by looking at a scatterplot. Once a design becomes more complicated then the existence of a collinearity will be more difficult to detect.

The purpose of this study is to investigate the effects of a collinearity in an analysis of covariance model and the importance of a testing strategy. This report is divided into two parts. The first part discusses the effects of a collinearity between covariate and treatment. These effects are examined under different tests of hypothesis, assuming different true models. The second part concerns the diagnostic procedures employed to detect a collinearity in analysis of covariance.

CHAPTER 2

Effects of Collinearity in an Analysis of Covariance Model

2.1 Statement of the Problem

In the standard use of the analysis of covariance model, the assumption is made that the range of values for the covariate is the same for all levels of the treatment. When this assumption is violated, the tests of hypothesis in a covariance analysis are affected and the F-tests may likely lead to the wrong conclusion concerning the presence of treatment effects. This problem could be described as a collinearity between the covariate and treatment, where collinearity is a term normally used in regression analysis. A broad definition of this term is that, if two variables lie "almost" on the same line, then they are collinear (Belsley, Kuh, and Welsch). By this definition, the covariate and treatment are collinear if the values of the X's are similar to each other within treatment levels but are dissimilar between treatment levels.

The extent of this problem is determined by the severity of the collinearity, the underlying true population model, and the testing strategy employed. How these factors influence the problem is the subject of this chapter.

In addition to determining the effects of the collinearity, there is also the problem of detecting the collinearity. Depending on the complexity of the analysis of covariance model, this may either be a simple, or a difficult task. If the model contains multiple covariates and/or multiple factors in the treatment structure it is possible that more than one collinearity may exist. In Chapter 3, methods of detecting and assessing