

INSTITUTE TEKNOLOGI MARA

MULTICOLLINEARITY IN REGRESSION ANALYSIS:  
EFFECTS, DETECTION AND METHODS OF MINIMISING.

Projek Diajukan Kepada  
Kajian Sains Matematik dan Komputer  
Sebagai Memenuhi Sebahagian Syarat  
Untuk Diploma Lanjutan Statistik

Oleh

AZMI BIN ABDUL LATIFF, Dip. Stats.

SHAH ALAM, SELANGOR

JUN, 1982

<u>CONTENT</u>	<u>PAGE</u>
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF APPENDICES	vii
CHAPTER 1 INTRODUCTION AND OBJECTIVES	1
CHAPTER 2 EFFECTS OF MULTICOLLINEARITY	6
ILLUSTRATION ON THE EFFECTS OF MULTICOLLINEARITY	9
Example showing the effect of multicollinearity	10
Effect of multicollinearity on regression sum of squares	12
CHAPTER 3 DETECTION OF MULTICOLLINEARITY	14
METHODS FOR DETECTING MULTICOLLINEARITY	15
The Analysis of Variance Table	15
Examining Correlation Coefficients of the Independent Variables	16
The Ridge Method	16
The Farrar-Glauber Test for Multicollinearity	18
CHAPTER 4 MEASURES TAKEN ON MULTICOLLINEARITY	26
METHODS USED IN MINIMISING MULTICOLLINEARITY	28
Reparameterization to Improve Computational Accuracy	28
The Ridge Method	38
Centering the Data Approach	45
CHAPTER 5 SUMMARY AND CONCLUSION	55
REFERENCES	

### ACKNOWLEDGEMENTS

I am most grateful to Puan Rasimah Aripin for her supervision, assistance and continued interest in the completion of this project. I would like to extend my gratitude to all lecturers and colleagues in the School of Mathematical Sciences and Computing for their cooperation, assistance and encouragement.

### LIST OF TABLES

<u>TABLE</u>	<u>TITLE OF TABLE</u>	<u>PAGE</u>
1	ANALYSIS OF VARIANCE TABLE FOR DATA SET 1 WITH DEPENDENT VARIABLE Y AND INDEPENDENT VARIABLES $X_1$ AND $X_2$ .	10
2	ANALYSIS OF VARIANCE TABLE FOR DATA SET 1 WITH DEPENDENT VARIABLE Y AND INDEPENDENT VARIABLE $X_1$ .	11
3	ANALYSIS OF VARIANCE TABLE FOR DATA SET 1 WITH DEPENDENT VARIABLE Y AND INDEPENDENT VARIABLE $X_2$ .	12
4	TRANSFORMED DATA FOR DATA SET 1.	33
5	THE RIDGE REGRESSION COEFFICIENTS VALUES FOR SUCCESSIVE VALUES OF K.	41
6	THE VARIANCE INFLATION FACTOR (VIF) VALUES FOR EACH REGRESSION COEFFICIENT AT SUCCESSIVE VALUES OF K.	42

## CHAPTER 1

### INTRODUCTION AND OBJECTIVES

The general purpose of this paper is to examine the effects of multicollinearity in regression analysis, explore ways of detecting and present methods of minimising these effects. The specific objectives of this paper are:

- (1) To examine the effect of multicollinearity;
- (2) To study and analyse the best possible methods used in detecting the presence of multicollinearity; and
- (3) To examine the possible methods that can best be used in minimising or reducing the effects of multicollinearity.

There are many approaches in detecting the presence of multicollinearity but the emphasis is more on suitable ones which consists of:

- (a) The Analysis of Variance Table (ANOVA Table);
- (b) Examining the correlation coefficients of the independent variable;
- (c) The Ridge method; and
- (d) The Farrar-Glauber Test.

The methods used to minimise the effects of multicollinearity are:

- (a) Reparameterization to improve computational accuracy;
- (b) Ridge Regression; and
- (c) Centering the input data.

The method of Reparameterization and Ridge Regression are useful in multiple regression while the method of centering