A COMPARISON OF METHODS FOR ANALYSING CORRELATED COUNT DATA.

By

Clair Alston

A THESIS SUBMITTED FOR THE DEGREE OF

MASTER OF SCIENCE

OF

THE UNIVERSITY OF NEW ENGLAND

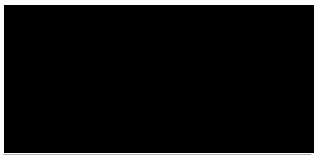
MARCH 1997

Preface

I hereby declare that this thesis describes my own original work, supervised by Dr Robert Murison (major supervisor), Dr David Smith and Dr Ian Davies.

I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree.

I certify that any help received in preparing this thesis, and all sources used, have been acknowledged in this thesis.



Clair Alston.

Acknowledgements

I would like to thank my supervisors, Dr Robert Murison and Dr David Smith, for their guidance and support with this work.

Graham Charles of NSW Agriculture, Narrabri, is thanked for providing the herbicide data set which I have used extensively in this thesis.

Thanks are also due to Adrian Doss who scribbled all over several "final" drafts with red ink, to Steven Harder for various computing tips, and to Dr Ian Davies, Tony and Barbara Bernardi for helpful comments on a recent draft.

Family and friends are also gratefully acknowledged for encouraging and tolerating me during the time I have spent doing this project.

Abstract

This thesis considers extensions of Generalized Linear Models (Nelder and Wedderburn, 1972) to incorporate correlated count data. Of particular interest is the Poisson random effects model which is commonly solved by approximate methods due to the complexity of calculations in maximum likelihood estimation (Diggle, Liang and Zeger, 1994, p173-5).

The methods considered fall into 4 categories;

1. quasi-likelihood techniques, (Schall, 1991), (Breslow and Clayton, 1993),
2. overdispersion models, (Van de Ven and Weber, 1995)
3. generalized estimating equations, (Liang and Zeger, 1986), and
4. Markov Chain Monte Carlo techniques, (Zeger and Karim, 1991).

These techniques are examined and compared both algebraically and through the use of a small simulation study. On this basis, some recommendations for the use of these methods in practice are made.

The variogram is used to determine which error model is appropriate to use with

a number of data sets, and use of several residual types resulting from GLMs are compared. This comparison is done so that the appropriate error model is most evident at the investigative stage of the analysis.

Contents

List of Tables

List of Figures