

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

THE ESTIMATION OF GENETIC PARAMETERS

FOR CATEGORICAL TRAITS

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Animal Science
at Massey University.

ARTHUR RICHARD GILMOUR

1983

18170-89

ABSTRACT

The estimation of heritabilities, genetic and environmental variances and covariances, and the prediction of breeding values, is a major concern among animal breeders. This study adapts existing statistical methods to provide a new method of estimating these parameters for categorical traits.

The problems associated with the analysis of categorical data arise because of the relationship between the mean and variance of the sampling distribution. The parameters of the sampling distribution are assumed to be a non-linear function of values on an underlying scale. It is further assumed that fixed and random effects are additive on the underlying scale. This scale cannot be observed and information about it must be deduced from the observed categorical trait.

Justify?
A common practice has been to estimate parameters (fixed effects and variances) on the categorical variable itself and then to transform these estimates to values applicable to the underlying scale. This procedure is theoretically invalid except for a fully random model in which the only fixed effect is the mean. The method developed in this thesis attempts to estimate parameters directly in the underlying scale by transforming the data before calculating estimates of the parameters. It is a synthesis of mixed model procedures (Henderson et al., 1959) and generalized linear models (Nelder and Wedderburn, 1972) and is called the generalized linear mixed model. The general method is for analysing data presumed to arise from a two-stage sampling procedure when the second sampling has an error distribution belonging to the single parameter exponential family.

Means

Justify
The detailed algebra for applying the new method to binomial and multinomial traits for the estimation of fixed effects is presented. The logit transformation is used in this application and the resulting system of equations is called the logistic linear mixed model. A procedure for estimating

variances, covariances and the intraclass correlation on the underlying scale is also developed.

The logistic linear mixed model is evaluated by comparing parameter estimates from the method with true values used to generate the data being analysed. Biases appear to be small except for some extreme combinations of parameters when assumptions made while developing the algebra break down.

The logistic linear mixed model is applied to two real problems for which fixed and random effects and variance components are estimated and comparisons made with parameters estimated by other methods. The first problem is the analysis of data on the feet characteristics of 2513 lambs, the second is the analysis of 1396 lambing performance records.

The concluding discussion considers the general use of the logistic linear mixed model and its relationship with other models.

ACKNOWLEDGEMENTS

I gratefully acknowledge the guidance of my supervisors, Professor R.D.Anderson, Professor A.L. Rae and Dr.G.A. Wickham.

The Australian Wool Corporation provided the post-graduate scholarship which enabled this study to be undertaken. It is thanked, along with the New South Wales Department of Agriculture which provided study leave and financial assistance.

The Animal Science Department has provided a fine working environment and the opportunity to broaden my outlook.

My gratiude is extended to the friends we have made, particularly in Milson. Their willingness to include us in their community has made life both pleasant and rewarding for me and especially for my wife, Ellen, and our three daughters. In particular, I acknowledge the friendship of Graham and Lynette Toms and the members of Milson Christian Fellowship.

To Ellen and the three girls, thank you for being there when needed and bearing with me when times were busy.

To my God, thank you for providing all our needs - friends, a home, work and recreation, health, finance and the opportunity to serve - through Jesus Christ.

Table of Contents

Section	Heading	page
	Abstract	ii
	Acknowledgments	iv
	Table of Contents	v
	List of Tables	ix
	List of Figures	xiv
Chapter 1.	Introduction	1
Chapter 2.	Literature review	3
2.1.	Categorical data	4
2.2.	Analysis on the binary (0,1) scale	6
2.2.1.	Binomial traits	6
2.2.2.	Multinomial traits	11
2.2.3.	Correlation between variables	13
2.2.4.	Selection involving categorical traits	15
2.3.	Analysis on the underlying scale	17
2.3.1.	Falconer's liability method	17
2.3.2.	Generalized linear models and other proposals	19
2.4.	Scope of the thesis	21
Chapter 3.	Review of statistical procedures required in later chapters	22
3.1.	Some statistical procedures used in animal breeding	22
3.1.1.	The selection index	22
3.1.2.	Best linear unbiased prediction	23
3.1.3.	Estimation of variance components	24
3.2.	Other statistical procedures	26
3.2.1.	Logistic distribution	26
3.2.2.	Log likelihood expectations	28

3.2.3.	Generalized linear models (GLM)	30
3.2.3.1.	Derivation of the general expressions for members of the single parameter exponential family	30
3.2.3.2.	Testing the 'goodness of fit' of a proposed model	35
3.2.4.	Maximum likelihood solution for parameters of the multinomial distribution by generalized linear models	35
3.2.5.	Solution of re-weighted least squares equations	39
Chapter 4.	Generalized linear mixed model	40
4.1.	The joint-maximization method	41
4.2.	The maximization-expectation method	44
Chapter 5.	Deriving the logistic linear mixed model	48
5.1.	Mixed model analysis of binomial data	48
5.1.1.	Solution by joint-maximization method	49
5.1.2.	Solution by the maximization-expectation method	49
5.1.2.1.	Using the Normal distribution	50
5.1.2.2.	Using the moments of a symmetric distribution	52
5.1.2.3.	Using the logistic distribution	55
5.2.	Analysis of multinomial data	59
5.2.1.	Analysis of extremal characters	59
5.2.1.1.	Solution by joint-maximization method	59
5.2.1.2.	Solution by the maximization-expectation method	60
5.2.2.	Analysis of multiple threshold characters	66
5.2.2.1.	Fixed effects generalized linear model	68
5.2.2.2.	Mixed model threshold analysis by joint-maximization method	70
5.2.2.3.	Mixed model threshold analysis by maximization-expectation method	71

Chapter 6.	Application of the logistic linear mixed model to animal breeding	75
6.1.	The simultaneous analysis of categorical and continuous characters by the same model	75
6.1.1.	Covariance between a binomial character and a normal character	77
6.2.	Estimation of variances	79
6.3.	Implementation of the logistic linear mixed model in a generalized linear models programme	83
6.3.1.	Measurement of 'lack of fit'	83
6.3.2.	Estimation of variance components	84
✓ Chapter 7.	Performance of the logistic linear mixed model in simulation studies.	87
7.1.	The behaviour of estimates of mean and variance under the logistic linear mixed model	87
7.2.	The precision of estimates of the threshold and the intraclass correlation	92
7.3.	The estimation of breeding values under the logistic linear mixed model	97
✓ Chapter 8.	A study of foot ailments associated with Merino-cross sheep grazing damp conditions	110
8.1.	Repeatability of assessment of foot-shape	111
8.2.	Estimation of heritability of incidence of foot ailments	112
✓ Chapter 9.	Reproductive performance of Perendale ewes	125
9.1.	Analysis of Perendale data by ordinary multivariate least squares	129
9.2.	Analysis of binary traits on the probit scale under a fixed effects model	130
9.3.	Analysis of the Perendale data using the logistic linear mixed model	135

Chapter 10. General Discussion	147
10.1. How relevant is an estimate of heritability on the underlying scale?	147
10.2. What relationship exists between the logistic linear mixed model and the ordinary mixed model analysis?	147
10.3. What can be done in very large problems where the iterative procedures would be prohibitively expensive?	149
10.4. When might it be more efficient to use equations based on the joint-maximization method rather than those based on the maximization-expectation method?	150
10.5. What further work is required?	152
10.6. Conclusion	153
Appendix A	
Some matrix symbols and operations	154
Appendix B	
More expectations for section 5.1.2.1 using the normal distribution	156
Appendix C	
Derivation of the basic expressions required to apply the generalized linear mixed model to multinomial data using the multinomial logit	159
Appendix D	
Some additional tables relating to chapter 8.	176
Bibliography	181

List of Tables

Table	page
3.1 The relationship between the standard normal and standard logistic distributions when used as probability transformations.	28
7.1 Effects of differing family size and numbers of families on estimates of the intraclass correlation and the threshold in binomial samples with extra-binomial variation.	94
7.2 Statistics for 10 samples of 100 $N(0, .10)$ random variables used to define the 10 flocks.	98
8.1 Analysis of variance of foot-shape scores.	111
8.2 Analysis of deviance of foot-shape scores when analysed as a double-threshold trait with an intraclass correlation of 0.731.	111
8.3 Iterative sequence for estimating the intraclass correlation for the double-threshold trait, L54 (lamb foot shape score 5, 4 and less than 4).	117
8.4 Analysis of deviance for the double-threshold trait L54 (lamb foot-shape score 5, 4 and less than 4).	117
8.5 Estimates of intraclass correlation and heritability (4 times intraclass correlation) for 12 sheep foot traits.	118

Table	page
8.6 Correlation between estimates of breeding value for foot-shape score, obtained by various methods.	118
8.7 Correlation between estimates of breeding value of 28 sires in four mating-groups for presence of foot-scald and presence of foot-rot.	119
8.8 Analysis of deviance for foot-shape traits L54 and H54, using 34 sires.	119
8.9 Analysis of deviance for foot-shape traits L5 and H5, using 34 sires.	120
8.10 Analysis of deviance for presence of foot-scald (LS) and presence of foot-rot (LR) in lambs using 28 sires.	120
8.11 Analysis of Variance for foot-shape score (LC) using 34 sires. K coefficient is 73.4388.	121
8.12 Analysis of variance for foot-shape trait L5 using 34 sires. K coefficient is 73.4388.	122
8.13 Analysis of Variance for presence of foot-scald (LS) using 28 sires. K coefficient is 69.7024.	123
8.14 Analysis of Variance for presence of foot-rot (LR) using 28 sires. K coefficient is 69.7024.	124
8.15 Incidence of foot-scald in lambs (5 months) on percentage and logit scales.	124

Table	page
9.1 Summary of the number of daughters sired by each of the 63 rams used in the flock from 1961 to 1972.	126
9.2 Dam age means for weaning weight, hogget weight and number of lambs weaned.	126
9.3 Year means for weaning weight, hogget weight and number of lambs weaned.	127
9.4 Variance and covariance components estimated by Henderson's (1953) method 3. Covariance components are below and correlations are above the diagonal	129
9.5 Heritabilities of five traits by Henderson's (1953) method 3.	129
9.6 Heritability on the probit scale by Falconer's (1965) method.	132
9.7 Heritability on the probit scale using Falconer's method after including hogget weight as a covariate and assuming genetic correlations of 1.	133
9.8 Sires in order of average rank for lambed-or-not and twins-or-not.	134
9.9 Fixed effects for all traits estimated by maximum likelihood single trait methods.	139

Table	page
9.10 Genetic and environmental variances, heritabilities and breeding values estimated by maximum likelihood single trait methods.	140
9.10 continued. Breeding values for all traits estimated by maximum likelihood single trait methods.	141
9.11 Some results from using various trial values for the variances and covariances associated with an extremal trait.	142
9.12 Genetic variances and covariances between two binary traits, lambed-or-not and twins-or-not, and hogget weight estimated for three environmental correlations, 0.0, 0.3 and -0.3.	143
9.13 Genetic variances and covariances between two categorical traits, lamb-t and lamb-e, and hogget weight, estimated for three environmental correlations, 0.0, 0.3 and -0.3.	144
9.14 Comparison of breeding values for lambed-or-not and hogget weight obtained assuming environmental correlation of 0.0 and genetic correlations of 0.0 and 0.6275.	145
9.15 Correlations between breeding values in table 9.14 and the rearing rank of the sire.	146
10.1 The probability that a group of nominated size and nominated incidence has observations all in the same category.	152

Table	page
D.1 Foot-shape scores recorded at two times by two observers on 97 lambs.	176
D.2 Foot-scald scores recorded at two times by two observers on 97 lambs.	176
D.3 Foot-rot scores recorded at two times by two observers on 97 lambs.	177
D.4a Experiment data for 1980 lambing	177
D.4b Experiment data for 1981 lambing	178
D.5 Estimates of breeding values for the trait foot-shape score - obtained by various methods.	179
D.6 Estimates of breeding values for the traits presence of foot-scald and presence of foot-rot, obtained by two methods.	180

List of Figures

Figure	page
7.1 Comparison of the scale factors associated with the three expressions for $E[p]$.	90
7.2 Comparison of the scale factors associated with the three expressions for $E[p(1-p)]$.	91
7.3 Relationship between actual and estimated values of the intraclass correlation coefficient for various threshold values.	95
7.4 Relationship between actual and estimated values of the threshold for various intraclass correlations.	96
7.5 Relationship between $B(.50)$ and family size.	103
7.6 Relationship between $G(.50)$ and family size.	103
7.7 Relationship between $B(.10)$ and family size.	104
7.8 Relationship between $G(.10)$ and family size.	104
7.9 Relationship between $B(.02)$ and family size.	105
7.10 Relationship between $G(.02)$ and family size.	105

Figure	page
7.11 Relationship between correlation (u, u_x) and family size.	106
7.12 Relationship between correlation (u, u_p) and family size.	107
7.13 Relationship between correlation (u_x, u_p) and family size.	108
7.14 Relationship between deviance (499 df) and family size.	109