

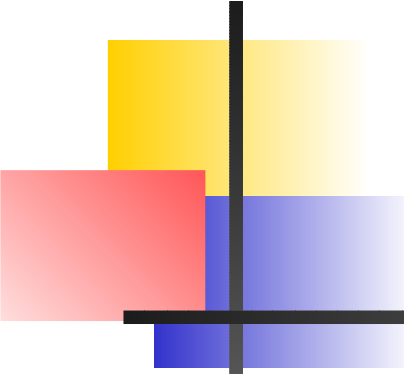
ASReml workshop

4.1 GLM & GLMM

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



Generalized Linear Mixed models

- Iteratively reweighted least squares
- Data types in exponential family
 - Binary $[0,1]$
 - Binomial $[r,n]$
 - Poisson
 - Gamma
 - Negative Binomial
- Variance is dependent on the mean

Method

- Analyse on a transformed scale using Link function and distribution specific weights

$$y = X\tau + Zu + D^{-1}(O - E)$$

where D is derivative of link function

O is the observed value

and E is $g(X\tau + Zu)$, the expected value

weights are a function of D and E

Method continued

- Apply usual REML to this working variable y updating E and D , weights and y each iteration.

- Binomial Logit

$$Y = X\hat{\tau} + Z\tilde{u}$$

$$\hat{E} = p = 1/(1 + e^{-Y}) \rightarrow Y = \log(p/(1 - p))$$

Derivative is $p(1 - p)$

Binomial weight is $np(1 - p)$

Analysis weight is $n/[p(1 - p)]$



Disclaimers

- Without random effects, this method is standard GLM (Generalized linear model) and ASReml also calculates the $-2L$, the used to assess improvement of fit from adding an extra term to the model (Analysis of Deviance).
- With random effects, the 'Deviance' can no longer be used to test changes in the fixed model, and the REML LogL reported cannot be used to test changes to the random model because y , the working variable changes between models



Families and Links

- !BIN [!ID !LOGIT !PROBIT !COMP] !TOTAL n
!POIS [!ID !LOG !SQRT]
!GAMMA [!ID !LOG !INV] [!phi p]
!NEGBIN [!ID !LOG !INV] [!PHI p]
General: !OFFSET o !DISP !DEV !WORK
!RESP !PEAR



Sire model

- Binomial has reasonable base for genetics because the logit [probit] link functions imply underlying residual logistic [normal] distributions with error variance 3.3 [1]
- Can fit the model for Poisson but I am unaware of proper genetic basis for estimating heritability say.

LAMB data [manual 9.8]

- Foot shape Score

```
# Yr Grp Sex Sire xxx tot l5 l4 ls
# 1 1 1 1 18 39 33 6 6 1
# 1 1 0 1 18 50 41 9 2 0
# 1 1 1 2 18 35 30 4 1 0
yr 2 Grp 5 Sex Sire 18
xxx tot L5 !/tot
L4 !/tot LS !/tot LR !/tot
lamb.dat !SKIP 1
L5 !BIN !LOG !TOTAL tot ~ mu
Grp Sex Sex.Grp !r Sire
```




ASReml workshop

4.2 QTL

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



Simple QTL

- Thredbo talk



Complex pedigrees

- IBD matrices

ASReml workshop

Further help

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



Internet

- support@vsn-intl.co.uk
- <http://www.asreml.co.uk/>
- ASReml discussion group
ASREML-L@agric.nsw.gov.au
To join,
<mailto:arthur.gilmour@agric.nsw.gov.au>
- Cookbook:
http://uncronopio.org/luis/asreml_co

ASReml workshop

4.3 ANOVA

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



The issues of ANOVA in REML

- Types of Sums of Squares
- Test statistic
- Denominator Degrees of Freedom

OATS - fixed ANOVA

- !TYPEIISS

`Y ~ mu bl var bl.wp nit nit.var`
ordered to get correct SS and DF.

BUT ASReml uses wrong F ratio because we have not specified that bl and wp are random.

- | Degrees | Freedom | Stratum | Variances | |
|----------|----------|---------|-----------|------|
| | 45.00 | 177.083 | 1.0 | |
| Source | Modterms | Gam | Component | C/SE |
| Variance | 72 45 | 1.0 | 177.083 | 4.74 |

Fixed ANOVA

AnOVar	Num	DenDF	F-incr	F-adj	K	Prob
mu	1	45.0	4395.31	NA	.	NA
blocks	5	45.0	17.93	17.93	A	<.001
variety	2	45.0	5.04	5.04	a	0.011
bl.wp	10	45.0	3.40	3.40	B	0.002
nitrogen	3	45.0	37.69	37.69	A	<.001
nit.var	6	45.0	0.30	0.30	B	0.932

Proper tests

- mu: $F = 245.14 = 4395.31 / 17.31$ with 1,5 DF
 - var: $F = 1.49 = 5.04 / 3.40$ with 2,10 DF
 - mu
- | | | | | | | |
|---------|----|------|---------|-------|---|-------|
| mu | 1 | 45.0 | 4395.31 | NA | . | NA |
| blocks | 5 | 45.0 | 17.93 | 17.93 | A | <.001 |
| variety | 2 | 45.0 | 5.04 | 5.04 | a | 0.011 |
| bl.wp | 10 | 45.0 | 3.40 | 3.40 | B | 0.002 |

OATS - REML analysis

■ !TYPEIISS

```

Y ~ mu var nit nit.var !r bl bl.wp
Degrees Freedom and Stratum Variance
      5.00      3175.06      12.0      4.0      1.
     10.00      601.331       0.0      4.0      1.
     45.00      177.083       0.0      0.0      1.

```

```

Source Modterms Gamma Component /SE
blocks      6      6  1.21  214.477 1.27
bl.wp      18     18 0.599  106.062 1.56
Variance  72     60 1.00  177.083 4.74

```

OATS - REML

- Proper F-incr value
- F-adj to be renamed F-con
- here same as F-inc because a balanced design.

AnOVar	Num	DenDF	F-incr	F-adj	K	Pro
mu	1	5.0	245.14	NA	.	N
nitrogen	3	45.0	37.69	37.69	A	<.00
variety	2	10.0	1.49	1.49	A	0.27
nit.var	6	45.0	0.30	0.30	B	0.93



Wald tests

- F-incr based on Sum of squares explained by adding the term in the order specified divided by NumDF and σ^2
- In mixed model, not easy to calculate DenDF.
Not available in ASReml 1.1
Not available for large models in ASReml 1.62
Not necessarily the same for F-incr and F-adj
Under active development.



Order of fitting

- Main effects before interactions:
 - If you specify an interaction first, the main effects will typically disappear
 - If you put an interaction in the sparse section, it will sweep out μ , and higher order terms.

In practice

- Interpret the following ANOVA

Src	NDF	DDF	F _{inc}	F _{adj}
X1	1	45	1.	21.
X2	1	45	23.	1.
X3	1	45	10.	10.

In practice

- Interpret the following ANOVA

Src	NDF	DDF	F _{inc}	F _{adj}
X1	1	45	1.	21.
X2	1	45	23.	1.
X3	1	45	10.	10.

- X1+X2+X3 explains 34.
X1+X2 explains 24
X1+X3 explains 33
X2+X3 explains 13

In practice

- Interpret the following ANOVA

Src	NDF	DDF	Finc	Fadj	K	%P
Region	2	45	11.	21.	b	
Site	11	45	8.	1.	A	
Var	12	45	10.	10.	A	
Var.Site	99	45	5.	5.	B	

In practice

- Interpret the following ANOVA

Src	NDF	DDF	Finc	Fadj	K	%P
Region	2	45	11.	21.	b	
Site	11	45	8.	1.	A	
Var	12	45	10.	10.	A	
Var.Site	99	45	5.	5.	B	

- Sites are nested in Region so Region cannot be tested after Site (Var.Site).

Conditional F-tests



Term	F-inc	F-adj
A	A	A B,C,B.C
B	B A	B A,C,A.C
A.B	A.B A,B	A.B A,B,C,A.C,B.C
C	C A,B,A.B	C A,B,A.B
A.C	A.C A,B,A.B,C	A.C A,B,C,A.B,B.C
B.C	B.C A,B,A.B,C,A.C	B.C A,B,C,A.B,A.C
A.B.C	A.B.C A,B,A.B,C,A.C,B.C	A.B.C A,B,A.B,C,A.C

More complicated example

■ AnOVar	NumDF	DenDF	F-incr	F-adj	K	Prob
FD0	1	220.0	36.81	34.30	C	<.001
GreenDM	1	10.7	9.52	8.59	A	0.014
Pos	1	10.0	93.09	93.03	A	<.001
Green.Pos	1	10.0	0.57	0.04	B	0.847
HM	1	9.5	0.12	3.39	B	0.095
Pos.HM	1	10.0	9.40	9.88	C	0.010
SLen	1	420.8	0.94	0.82	A	0.368
SL.Pos	1	1004.2	1.66	1.72	B	0.190
SL.Green	1	396.3	4.18	6.18	B	0.014
SL.Pos.Green	1	1275.0	1.55	11.52	C	<.001

More complicated example

■ at (Group, 4)	1	11.2	5.02	0.02	A	0.878
at (Group, 1)	1	9.4	5.98	0.00	A	0.978
at (G, 4) .Pos	1	10.1	0.22	0.60	B	0.456
at (G, 1) .Pos	1	9.9	0.99	0.15	B	0.703
at (G, 4) .Pos .SL	1	1786.9	3.20	18.67	C	<.001
at (G, 1) .Pos .SL	1	1125.4	4.69	10.95	C	0.001
Clover	1	9.8	0.71	0.73	A	0.414
Clov .SL	1	535.4	1.07	1.07	B	0.302
Clov .Pos	1	10.0	0.57	0.57	B	0.467
Clov .SL .Pos	1	1269.5	16.06	16.06	C	<.001



ASReml workshop

Further help

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



Internet

- support@vsn-intl.co.uk
- <http://www.asreml.co.uk/>
- ASReml discussion group
ASREML-L@agric.nsw.gov.au
To join,
<mailto:arthur.gilmour@agric.nsw.gov.au>
- Cookbook:
http://uncronopio.org/luis/asreml_cookbook.html



ASReml workshop

Prediction in Linear Mixed Models

Arthur Gilmour



NSW DEPARTMENT OF
PRIMARY INDUSTRIES



Prediction

- This is the process of calculating linear combinations of the effects to summarise aspects of the analysis. e.g.
 - plotting a fitted spline curve - creating treatment means
- This presentation is based on two previous presentations Arthur Gilmour, Sue Welham, B Cullis, B Gogel & R Thompson



Plan

- Introduction - The linear mixed model
- Prediction in Large problems
- Estimability
- Random effects
- Specification



linear mixed model

- $y = X\tau + Zu + e;$
 $y \sim N(X\tau, R + ZGZ);$
 $u \sim N(0, G); e \sim N(0, R)$

Mixed model equations

- $$\begin{pmatrix} X R^{-1} X & X R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \tau \\ u \end{pmatrix} = \begin{pmatrix} X R^{-1} y \\ Z R^{-1} y \end{pmatrix}$$
 rewritten as $C\beta = W R^{-1} y$

giving $\beta = C^{-1} W R^{-1} y$ prediction $\pi = D\beta$
with variance $D C^{-1} D$

Large Models

- ASReml fits large models; avoids forming all elements of C^{-1}

- Absorbing C in
$$\begin{pmatrix} yR^{-1}y & 0 & yR^{-1}W \\ 0 & 0 & D \\ WR^{-1}y & D' & C \end{pmatrix}$$

gives
$$\begin{pmatrix} yPy & -\pi \\ -\pi & -DC^{-1}D' \end{pmatrix}$$

Estimability

- No data. Height of males, females and angels
- Overparameterization: model $\mu + \alpha_i$
 α_i is non-estimable - infinitely many solutions

μ	α_1	α_2	α_3
0	10	12	14
10	0	2	4
12	-2	0	2

Estimability

- Over-modelling $\mu + \alpha_i + \beta_j$

12	10	14	*	*
11	13	15	*	*
*	*	*	17	18



Averaging over incomplete tables

- No problem if table is all estimable even though some cells have no actual data I.e. interaction is omitted or is random
- Sometimes sensible to average only over cells containing data. E.g. variety.year.location table: form variety table averaging over the experiments which form an incomplete year.location table.

Detection of Nonestimability

- During absorption process, nonestimability is present when the row of D is not zero but the row of C is zero.

$$\begin{array}{cc} -25/5 & -1/5 \end{array}$$

$$\begin{array}{ccc} * & 0 & -1/0 \end{array}$$

$$\begin{array}{cccc} -30/6 & 0 & 0 & -1/5 \end{array}$$

$$\begin{array}{ccccc} 25 & 1 & 0 & 0 & 5 \end{array}$$

$$\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \end{array}$$

$$\begin{array}{ccccccc} 30 & 0 & 0 & 1 & 0 & 0 & 5 \end{array}$$



Random effects

- The residual term is usually ignored. But kriging is prediction based on a correlated residual.
- Other random terms might be
 - error terms (usually ignored) or
 - treatment terms - may be predicted, averaged (conditional) or ignored (marginal).
- averaging over random lowers SE, may not affect SED

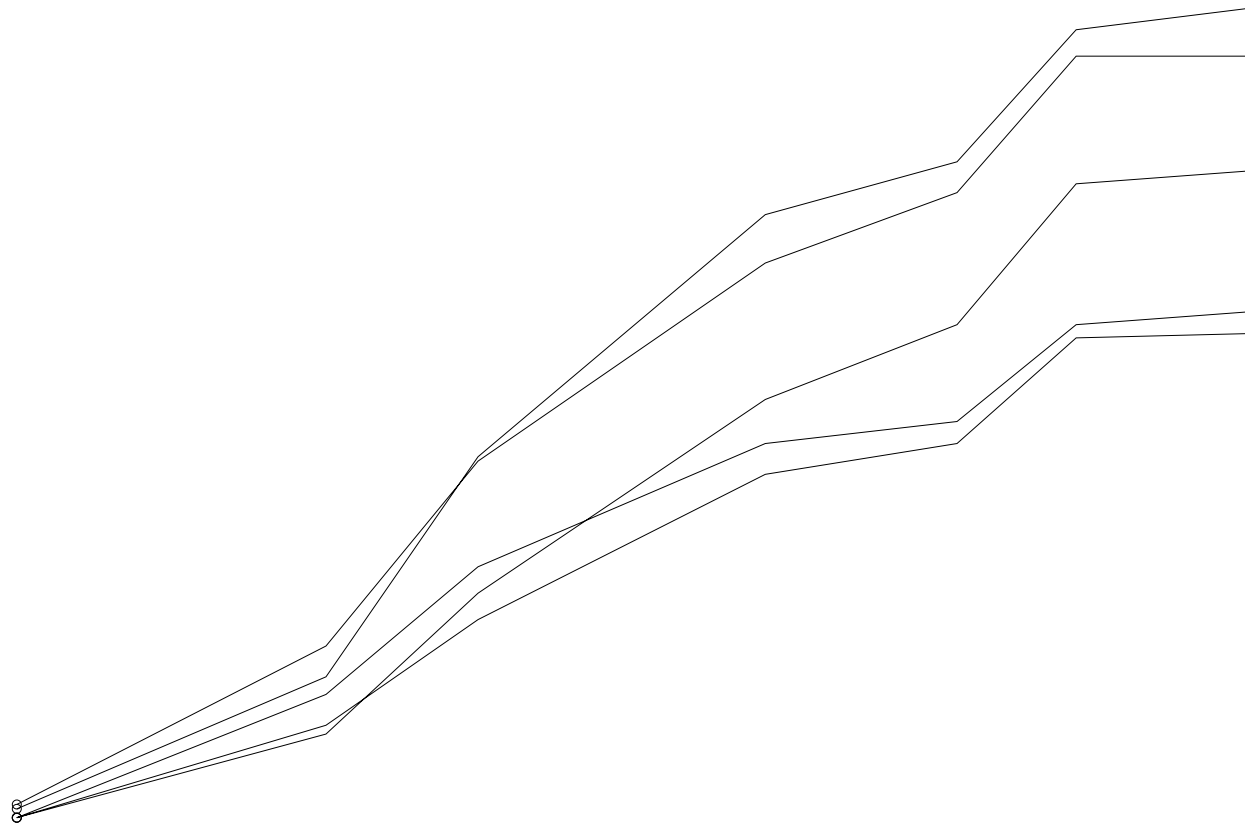


Specification

- Mixed model consists of model terms derived from factors and covariates.
 - Classify set - defines table to predict
 - Average set - usually other fixed variables
 - Ignored set - usually other random variables
- Set specific levels to be predicted
- Whether to average over cells with data
- Specific in/ex-clusion of model terms
- Whether to print nonestimable solutions

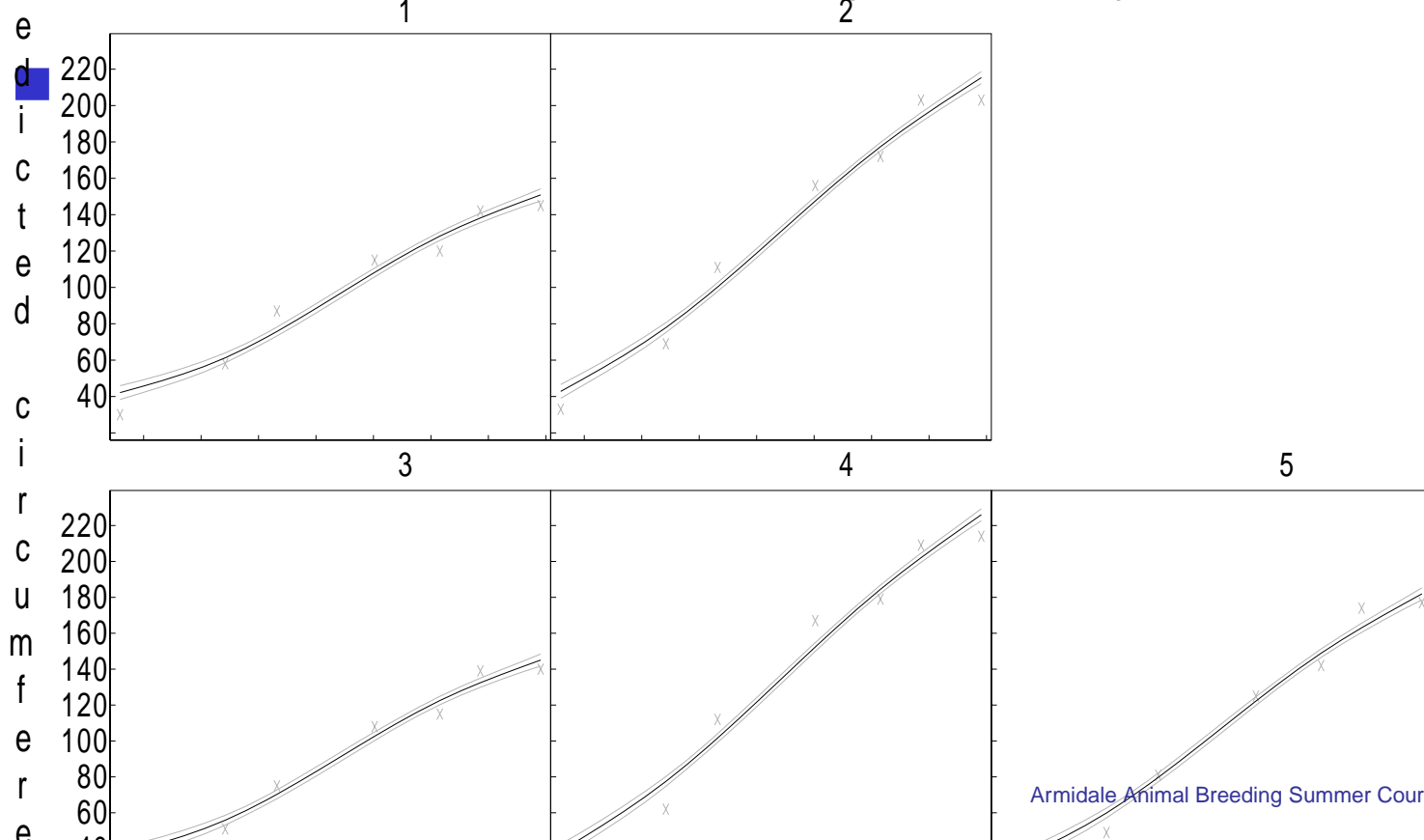
Growth of 5 Orange trees

Title: orange.
Y-axis: 30.0000 214.0000 Y=circumference X=age X-axis: 118.0000 1582.0000



Orange tree model

- Fixed - mu season day
- Random - Tree Tree day spl(day)





Conclusion A

- Ability to predict from a linear model is essential for reporting results
- User must be advised if predictions or non-estimable functions
- Algorithm must be able to handle large and complex models
- Algorithm must give control of how to handle random terms



Predict directive

- basic PREDICT syntax
- Where to place the PREDICT directive
- additional qualifiers
- other issues - spline interpolation



Basic syntax

- PREDICT classify set [qualifiers]
 - !present present set,
 - !average factor [weights]
 - Each factor name in the classify and present sets may be followed by list of actual values to predict



Where it goes

- Immediately after the linear model (before the Variance header line)
 - or after all R and G structure lines
- There can be many PREDICT statements.
 - Parsed at start
 - Design formed before first iteration
 - Prediction done in final iteration



Basic concepts

- Hyper table defined by all factors in the model
e.g. breed year animal sex
- Collapsed by either averaging or ignoring dimensions to produce the predict table.
- E.g $y = \mu + \text{variety} + r$ block predict variety



Hypertable control

- `!PARALLEL a b c`
 - In model need to link factor and covariate versions of same data with `lin()` or `fac()`
- data line qualifiers `!PPOINTS` and `!PVALS` for covariates (especially splines)
- specify particular levels of classify and present factors.
 - Predict breed sex 1

Fine control - which model terms to ignore

- Default - ignore model terms involving hyper factors which only occur in random model terms (e.g. animal but not spl())
- Control
 - !USE model_terms
 - !ONLYUSE model_terms
 - !IGNORE model_terms
 - !EXCEPT model_terms



Output control

- !PRINTALL
- !SED, !VPV
 - increased memory required for V matrix
- Backtransformation !LOGIT !PROBIT
!INVERSE !SQRT !COMPLOGLOG !LOGN
!POWER p Deprecated qualifier
- !FITMARGIN if classify set is two-way table
causes marginal tables to be produced.



General

- !FINAL command line option does one !CONTINUE iteration making it easy to do modelling then add predict statements at



Tabulate directive

- TABULATE <y-list> [!stats] <factors> e.g.
TABULATE Yield !count variety
- Enable simple investigation of the data
 - There may be several TABULATE statements;
 - Immediately after linear model
 - they are processed immediately
 - result written to .tab file