

Example 3 (referred to in module 4)

Restricted Maximum Likelihood Method to estimate components of genetic variation – an example in quantitative methods

John Rowlands

International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya

Background

Helminths constitute one of the most important constraints to livestock production in the tropics. Widespread infection with internal parasites in grazing animals, associated production losses, costs of anthelmintics and death of infected animals are some of the major concerns. Current control methods focus on reducing contamination of pastures through anthelmintic treatment and/or controlled grazing. In the tropics, these methods are limited by the high costs of anthelmintics, their uncertain availability and increasing frequency of drug resistance. In this situation an attractive, sustainable solution is breeding for disease resistance. Indeed, there is a large and diverse range of indigenous breeds of sheep and goats in the tropics, some of which appear to have the genetic ability to resist or tolerate helminthiasis. This was the background to a previous case study on the use of fixed effect least squares analysis to investigate the genetic resistance to helminthiasis of two indigenous breeds of sheep – Dorper and Red Maasai. As well as comparing resistance to helminthiasis between breeds it is also of interest to examine genetic variation within breeds. For this we need to use what are known as ‘restricted’ or ‘residual maximum likelihood’ (REML) procedures which are able to simultaneously estimate random and fixed effects and use the random estimates to determine heritability.

As for the previous case study the data used in this example comes from a study carried out at Diani Estate of the Baobab Farms, 20 km south of Mombasa in the sub-humid coastal region of Kenya between 1991 and 1997. The purpose of the experiment was to compare the genetic resistance to helminthiasis of two breeds of sheep – Dorper, which is indigenous to South Africa and Red Maasai, which is indigenous to Kenya. Throughout the six years Dorper (D), Red Maasai (R) and FI (RXD) ewes were mated to Red Maasai and Dorper rams to produce a number of different lamb genotypes. For the purposes of this example, only the following four offspring genotypes are considered:-

D x D, D x R, R x D and R x R. (For shorthand we shall use DD, DR, RD and RR with the first letter referring to sire and the second dam.) The lambs within these genotypes were born to 74 rams and 367 ewes. A total of 882 lambs resulted from the above genotype matings. Thus, each ewe gave birth on average to more than two lambs, one each in a different year.

Measurements of body weight were made periodically over a period of about a year on these lambs. Other characteristics recorded included weaning age, date of birth, the lamb’s sex, the age of its dam and the identity of lamb’s sire and dam. As in the previous example we shall consider weaning weight as the response variable.

Least squares analysis of variance

Before using the REML method to estimate genetic variance components we shall rerun the least squares analysis using a slightly different model to that used in the previous study to compare the genetic resistance of Red Maasai and Dorper sheep. This time we shall alter the way that breed genotypes are defined. Instead of referring to the breeds by their genotype, D X D, D X R, R X D and R X R we shall consider the breed of sire, the breed of dam and their interaction, and re-parameterise the model accordingly. The output looks as follows where:

WEANWT is the response variable (weaning weight),
 YEAR is the year of birth of the lamb,
 SEX is the sex of the lamb,
 AGEWEAN is the weaning age of the lamb,
 DL is the linear term for age of dam,
 DQ is the quadratic term for age of term,
 DAM_BRD and SIRE_BRD are the breeds of dam and sire,
 respectively.

Regression Analysis

Response variate: WEANWT

Fitted terms: Constant + YEAR + SEX + AGEWEAN + DL + DQ +
 SIRE_BRD + DAM_BRD + SIRE_BRD.DAM_BRD

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ YEAR	5	1208.149	241.630	48.92	<.001
+ SEX	1	55.983	55.983	11.34	<.001
+ AGEWEAN	1	344.206	344.206	69.69	<.001
+ DL	1	151.513	151.513	30.68	<.001
+ DQ	1	275.795	275.795	55.84	<.001
+ SIRE_BRD	1	44.881	44.881	9.09	0.003
+ DAM_BRD	1	30.223	30.223	6.12	0.014
+ SIRE_BRD.DAM_BRD	1	0.754	0.754	0.15	0.696
Residual	687	3392.947	4.939		
Total	699	5504.450	7.875		

Because the interaction was not significant, the model was run again to determine parameter estimates for breed not influenced by the interaction term.

Estimates of parameters

	estimate	s.e.	t(688)	t pr.
Constant	2.731	0.925	2.95	0.003
YEAR 92	-1.566	0.293	-5.35	<.001
YEAR 93	-1.096	0.275	-3.98	<.001
YEAR 94	-2.833	0.358	-7.92	<.001
YEAR 95	-3.228	0.344	-9.39	<.001
YEAR 95	-3.228	0.344	-9.39	<.001
YEAR 96	-2.351	0.390	-6.03	<.001
SEX M	0.478	0.169	2.82	0.005
AGEWEAN	0.07022	0.00886	7.93	<.001
DL	2.189	0.249	8.80	<.001
DQ	-0.2689	0.0340	-7.91	<.001
SIRE BRD R	-0.443	0.173	-2.56	0.011
DAM BRD R	-0.586	0.237	-2.48	0.014

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ YEAR	5	1208.149	241.630	48.99	<.001
+ SEX	1	55.983	55.983	11.35	<.001
+ AGEWEAN	1	344.206	344.206	69.78	<.001
+ DL	1	151.513	151.513	30.72	<.001
+ DQ	1	275.795	275.795	55.91	<.001
+ SIRE BRD	1	44.881	44.881	9.10	0.003
+ DAM BRD	1	30.223	30.223	6.13	0.014
Residual	688	3393.701	4.933		
Total	699	5504.450	7.875		

Before studying the magnitude of the sire and dam variance components, it is informative to compare the output obtained by running the same model through the REML procedure with that achieved by least squares analysis of variance. The following shows the Genstat output when the REML procedure is used.

REML Variance Components Analysis

Response Variate : WEANWT

Fixed model :
 Constant+YEAR+SEX+AGEWEAN+DL+DQ+SIRE_BRD+DAM_BRD
 No random model specified

Number of units: 700

No absorbing factor

* Residual term has been added to model

* All covariates centred

Estimated Variance Components

Random term	Component	S.e.
units	4.933	0.266

Approximate stratum variances

units	4.933	Effective d.f. 688.00
---------	-------	--------------------------

Matrix of coefficients of components for each stratum

units 1.00

*** Deviance: -2*Log-Likelihood ***

Deviance	d.f.
1855.80	688

Note: deviance omits constants which depend on fixed model fitted.

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.
YEAR	244.9	5
SEX	11.3	1
AGEWEAN	69.8	1
DL	30.7	1
DQ	55.9	1
SIRE BRD	9.1	1
DAM BRD	6.1	1

* All Wald statistics are calculated ignoring terms fitted later in the model

*** Table of effects for Constant ***

1
12.95

Table has only one entry: standard error 0.2261

Table of effects for YEAR

YEAR	91	92	93	94	95	96
	0.000	-1.566	-1.096	-2.833	-3.228	-2.351

Standard error of differences	Average	0.3373
	Maximum	0.3898
	Minimum	0.2753
Average variance of differences		0.1151

Table of effects for SEX

SEX	F	M
	0.0000	0.4779
Standard error of differences:	0.1695	

Table of effects for AGEWEAN

1
0.07022

Table has only one entry: standard error 0.008856

Table of effects for DL

1
2.189

Table has only one entry: standard error 0.2488

Table of effects for DQ

1
-0.2689

Table has only one entry: standard error 0.03401

Table of effects for SIRE_BRD

SIRE_BRD	D	R
	0.0000	-0.4429
Standard error of differences:	0.1728	

Table of effects for DAM_BRD

DAM_BRD	D	R
	0.0000	-0.5855
Standard error of differences:	0.2366	

You will note that when no random term is specified REML and the least squares analysis obtains the same solutions. Just the format of the output is different. The parameter estimates and associated standard errors from both analyses are exactly the same. However, instead of F-values Genstat calculates what are known as Wald statistics. The Wald test investigates the same hypotheses as the F test in the split-plot analysis of variance – i.e. null hypothesis of no effect - but unlike the F-statistic, which follows an F-distribution, the Wald test statistics

follow a Chi-square distribution, but only approximately. Significance levels tend to be a little lower for the Wald test compared with the F-test, and this will, by and large, always be the case unless sample size, as in this example is large. To derive the corresponding F-value from the value of the Wald statistic one just needs to multiply the Wald statistic by the corresponding degrees of freedom. Thus, the F-value for the linear term DL for instance is 30.72 the same as the corresponding Wald statistic (namely 30.7). This is because both have one degree of freedom. Note also that the estimated residual variance is the same in both outputs i.e. $\sigma_e^2 = 4.933$ with 688 degrees of freedom.

The table for Wald tests shows that the main effects of sire breed and dam breed adjusted for sire breed are significant with Wald statistics of 9.1 and 6.1. These values are seen to be significant ($P < 0.01$ and $P < 0.05$, respectively) when looking up a chi-square standard table with 1 degree of freedom. These levels of significance are the same at the 0.01 and 0.05 level, respectively, as those shown alongside the F-values on the least squares analysis of variance.

The Wald statistics also demonstrate the highly significant fixed effects of year of birth, age of dam, age at weaning and to a lesser extent sex on weaning weight shown earlier by least squares analysis. The table of effects demonstrate that the lambs born in the earlier years had higher weaning weights than those born in the later years and further, that male lambs had an average weaning weight slightly higher by 0.48 (± 0.17) kg than females.

The main points to be aware of in the interpretation of the REML output are that:- (a) the validity of the Wald testing depends on the sample size being large enough and (b) that the tests are more liberal than the F-tests, with the significance levels of the two becoming more similar with increasing sample size. Other statistical packages may use F rather than Wald tests. Nevertheless, the same comments apply and the user needs to take care in calculating significance values for the F-tests in mixed model analysis.

Mixed models and data structure

Before incorporating sire and dam effects into our model it is worth digressing and discussing the meaning of mixed models. Mixed model methodology takes its name from the fact that the elements of the model underlying the analysis can be a mixture of what are called fixed and random effects. The methodology is also known as REML analysis (where REML stands for residual, or restricted, maximum likelihood). The approach has become important in the analysis of data that have a hierarchical structure, since the different layers in the structure can be modelled using random effects.

One fundamental step in using mixed modelling for hierarchical data, is recognising the structure – i.e. the different layers - of the data. In order to help with this we shall use what we describe as a “mixed model tree” to develop the different layers pictorially.

Defining fixed and random effects

A random effect is a component of the data which has a degree of randomness associated with it, whereas a fixed effect has no random connotation. Thus, an example of a fixed effect

would be the sex of the offspring lambs in this experiment. Whilst it is not an effect, which can be specifically designed into the study, we have nevertheless demonstrated that the males grew significantly faster than females. The sex of the lamb is a fixed effect because there is no element of randomness in its average effect on growth. On the other hand, the influence of the sire on the growth of its offspring in breeding experiments is usually a random effect, since the researcher is interested in quantifying the extent to which different sires selected at ‘random’ from a breed have a bearing on growth, as opposed to comparing individual sires within the experiment itself.

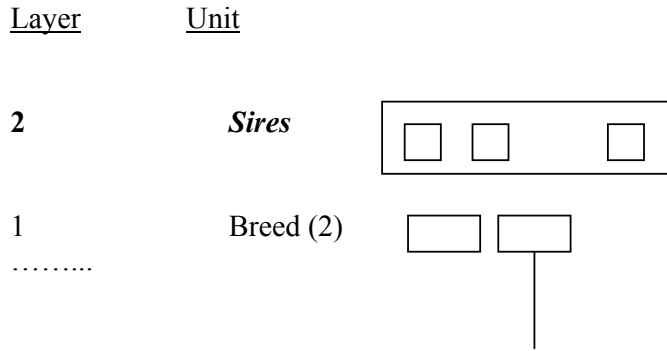
By defining an effect as random we are visualising the set of units as a representative sample from a much wider population. Thus, in this example, we regard the sires as a random representation of Red Maasai or Dorper ram breeds. If such an effect is defined as random then any interaction involving the effect and any other effect, fixed or random, will also be random – and this has implications for the inferences made from the data. For instance if year is declared random in this model, and the breed \times year interaction is included in the analysis, then any inferences made about breed will be for the population of years which our sample is deemed to represent. If the breed \times year interaction is not included, or if year is regarded as a fixed effect, then inferences will apply to the performance of the breeds only across the six years in question.

The choice of whether an effect such as breed is fixed or random is not always obvious. In this example there are only two breeds of sire and so it would not be sensible to infer that these two breeds are a random sample from a much larger population of breeds of ram. This is not only because they were two specifically chosen breeds, but also because a sample of two would not be considered large enough to generalise to “all breeds”. In the case where a study such as this is carried out over three or four consecutive years, one would need to consider whether that sample is large, or long, or random enough to be representative of the much wider population of time.

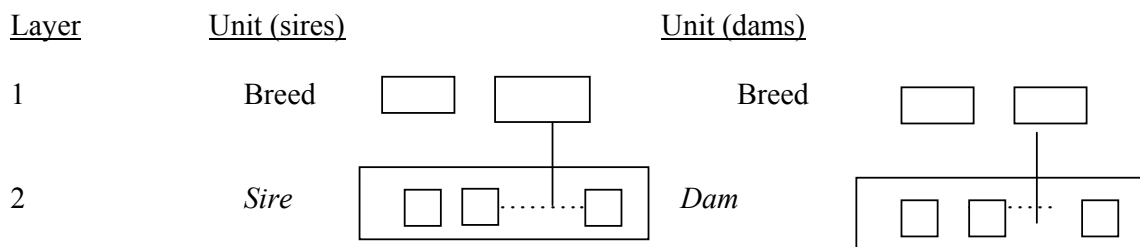
Units of investigation

In mixed model analysis we have different types of units occurring at the different layers - e.g. lambs within dams. The investigational units defined within a layer are assumed independent of one another; normally these are chosen at ‘random’ – they will therefore be random effects in our mixed model

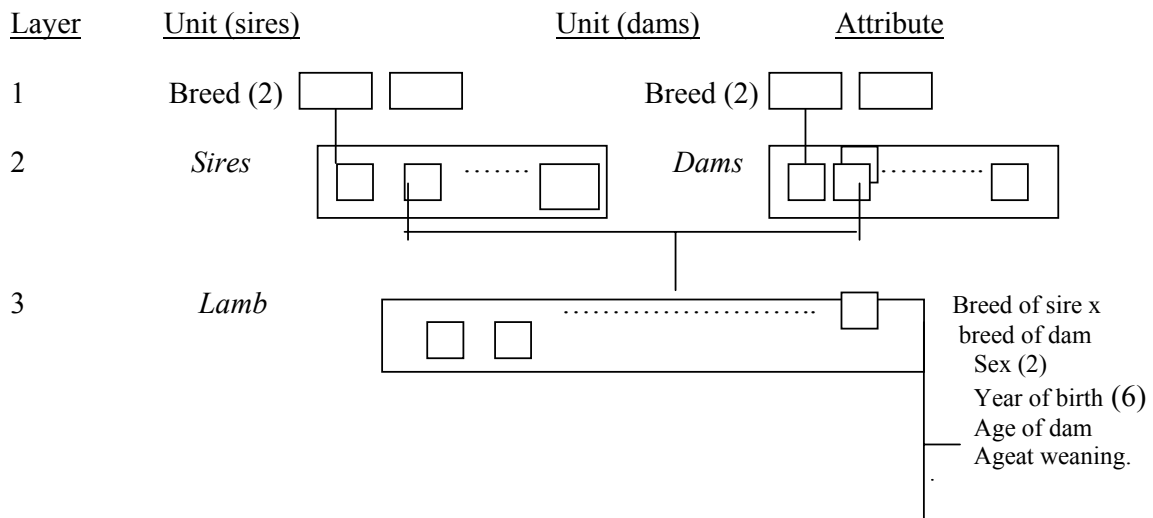
Correctly identifying the layers of units and the different attributes assigned to units is crucial to successful modelling of hierarchical data, and this is what we aim to do with our “mixed model tree”. Below is the hierarchical structure of the design in this study. Let us take sires first. The first layer describes the two sire breeds: Red Maasai and Dorper. From within each of the two breeds a number of sires is selected. These are the investigational units (shown in italics below to indicate a random effect) against which the two breeds - investigational units at the layer above - are compared.



We can do exactly the same for dams. As the selection process is being carried out at the same time as the sires, the mixed model tree is formed in parallel.



Sires and dams are mated both within and across breeds to produce their offspring. These offspring are the investigational units at the next layer down shown together with a list of attributes that might be considered for each lamb.



Whilst we have defined breed at the top layer, we will not consider it as an investigational unit because there is no replication. Instead we treat the breeds as fixed effects associated with the sire and dam investigational units. They are therefore assessed relative to the residual variation among sires and among dams. On the other hand the interaction between the breed of sire and the breed of dam is a feature individual to each offspring. The size of the interaction is, therefore, evaluated at the lowest level, i.e. relative to the variation among

lambs. The other fixed effects also are attributes pertaining to individual lambs and therefore feature at the lamb layer.

The REML Analysis

The following output demonstrates the mixed model analysis for lamb weaning weight. Sire and dam are defined as random effects; the remaining parameters as fixed effects. Note that since in this example each sire and dam is identified with its own unique code, Genstat detects that the random components are to be calculated within breed. Had sire and dam been coded within breed (e.g. as cumulative integers from 1 onwards) the random terms would have to be specified as SIRE_BRD.SIRE_ID and DAM_BRD.DAM_ID (interpreted as sire within sire breed and dam within dam breed respectively). In the following analysis the interaction term sire breed.dam breed is omitted as it was found to be not significant in the previous fixed model.

REML Variance Components Analysis

Response Variate : WEANWT

Fixed model: Constant+YEAR+SEX+AGEWEAN+DL+DQ+SIRE_BRD+DAM_BRD

Random model : SIRE_ID+DAM_ID

Number of units : 700

Estimated Variance Components

Random term	Component	S.e.
SIRE_ID	0.067	0.089
DAM_ID	1.457	0.283
units	3.427	0.266

Approximate stratum variances

	Effective	d.f.
SIRE_ID	4.733	57.66
DAM_ID	6.490	297.74
units	3.427	332.60

Matrix of coefficients of components for each stratum

SIRE_ID	10.31	0.42	1.00
DAM_ID	0.00	2.10	1.00
units	0.00	0.00	1.00

Deviance: -2*Log-Likelihood

Deviance	d.f.
1817.10	686

Note: deviance omits constants which depend on fixed model fitted.

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.
YEAR	230.3	5
SEX	9.7	1
AGEWEAN	63.8	1
DL	30.4	1
DQ	78.4	1
SIRE BRD	6.6	1
DAM BRD	2.9	1

* All Wald statistics are calculated ignoring terms fitted later in the model

Table of effects for Constant

1
12.87

Table has only one entry: standard error 0.2311

Table of effects for YEAR

YEAR	91	92	93	94	95	96
Effect	0.000	-1.571	-1.077	-3.003	-3.288	-2.450

Standard error of differences: Average 0.3226
Maximum 0.3947
Minimum 0.2590
Average variance of differences: 0.1056

*** Table of effects for SEX ***

SEX F M
0.0000 0.4038

Standard error of differences: 0.1623

*** Table of effects for AGEWEAN ***

1
0.06593

Table has only one entry: standard error 0.008613

*** Table of effects for DL ***

1
2.342

Table has only one entry: standard error 0.2331

*** Table of effects for DQ ***

1
-0.2900

Table has only one entry: standard error 0.03179

*** Table of effects for SIRE_BRD ***

SIRE_BRD	D	R
	0.0000	-0.4130
Standard error of differences:		0.1756

*** Table of effects for DAM_BRD ***

DAM_BRD	D	R
	0.0000	-0.4542
Standard error of differences:		0.2665

Genstat can also be requested to produce predicted means – these are equivalent to least square means in the analysis of variance.

*** Table of predicted means for Constant ***

1
10.74

Table has only one entry: standard error 0.2311

*** Table of predicted means for YEAR ***

YEAR	91	92	93	94	95	96
	12.61	11.07	11.56	9.64	9.35	10.19
Standard error of differences:			Average			0.3226
			Maximum			0.3947
			Minimum			0.2590
Average variance of differences:						0.1056

*** Table of predicted means for SEX ***

SEX	F	M
	10.54	10.94
Standard error of differences:	0.1623	

*** Table of predicted means for SIRE_BRD ***

SIRE_BRD	D	R
	10.95	10.54
Standard error of differences:	0.1756	

*** Table of predicted means for DAM_BRD ***

DAM_BRD	D	R
	10.96	10.51
Standard error of differences:	0.2665	

This output differs from the one given earlier in that the error variance is now shared among the specified random terms in the model. This ensures that different fixed estimates are evaluated against standard errors determined from residual variation associated with the appropriate layer(s). Note that compared with the earlier analysis the Wald statistics for sire breed and dam breed adjusted for sire breed are reduced; indeed the effect of the breed of dam is no longer significant. The standard error for comparing sire breed is now 0.176 kg compared with 0.173 kg in the previous analysis. This is because this standard error also takes into account the variation among sires within breeds. The standard error for comparing dam breed has increased even more – from 0.237 to 0.266 kg. This may be attributed to the fact that the variance component due to dam is much higher than that due to sire.

Another notable difference between this output and the one given earlier is the addition of two other terms under each of the headings; estimated variance components, approximate stratum variances and the matrix of coefficients of components for each stratum. This matrix of coefficients multiplied by the vector of estimated variance components gives the vector of approximate stratum variances. Variance components have an important use because they provide the basis for calculating genetic parameters such as heritability. In this analysis the dam variance component is much higher than the sire component indicating the large maternal influence on weaning weight of the lamb. The matrix of coefficients shows that the sire stratum variance includes a proportion (0.42) of the dam variance component. The 2.10 coefficient in the dam line indicates the average number of offspring per dam and since a sire was never mated to the same dam twice, the dam stratum variance is independent of sire (indicated by the 0.00 value for sire). With the random model specified the estimate of the residual among lambs variance is reduced from 4.939 to 3.427 kg, together with the effective degrees of freedom from 688 to 332.6. This is a result of taking out the variation among sires and dams within breeds. The earlier output assumes all variation to be at the lamb level. Therefore, the REML analysis with the random model describes more accurately the difficult layers of variation associated with hierarchical data and provides a more appropriate and correct analysis.

Comparison of the SIRE_ID and DAM_ID variance components with their standard errors indicates that the variance component for dam is highly significant (component > 5 times its standard error) although that for sire is not (component < its standard error). This therefore gives good evidence that the mixed model gives much more information contained in the data than the model without these two components.

Another way to compare models is to use what is known as ‘deviance’. In simple terms this gives an overall measure of how well the model fits the data. It is calculated as -2 times what is known as the log likelihood. While the method of least squares is usually adopted in fitting models involving fixed effects only, the method of maximum likelihood is the method used in REML. This method calculates an expression known as the likelihood, which measures how well the model fits the data. The better the fit the smaller the value of $-2\log\text{likelihood}$. By comparing the deviance values derived from separate models one can determine which model provides a better fit to the data. In this example the statistic can be used to determine whether the model that specifies the separate variance components fits the data better than the fixed model which contains only the residual variance component. The deviance in this output, namely 1817.10, is lower than that of the former, namely 1855.80. The difference between the two values $1855.80 - 1817.10 = 38.70$ with $688 - 686 = 2$ degrees of freedom. This approximates to a chi-square distribution ($\chi^2_2 = 38.70$). This is significant ($P < 0.001$) and shows that the mixed model provides a better fit.

A strong reason for using mixed modelling is that it can deal effectively with different layers in the data, it gives more valid significance tests and provides appropriate and correct standard errors – something that conventional analysis of variance methods cannot do except in one or two very specific circumstances. Secondly, it has the ability, with particularly unbalanced or sparse data structures, to combine information from different layers in the data. This has the advantage of improving the precision of the comparisons of fixed effects. These two reasons are why mixed modelling should be used in the analysis of some of the messy hierarchical structures that can occur in animal breeding research. Nevertheless, it may be sensible to evaluate some of the important fixed effects first and then add the random terms later as has been done here.

Genstat does not have the facility to calculate heritability estimates directly. An ASREML program developed by NSW Agriculture (1999) can do so. The ASREML program output is slightly different from Genstat. A part of the output is given below for the same data set:

ASREML Diani data

Source	Model	terms	Gamma	Component	Comp/SE	% C
Sire	5338	74	0.194259E-01	0.665780E-01	0.75	0 P
Dam	12682	358	0.424977	1.45651	5.13	0 P
Variance	700	688	1.00000	3.42727	12.83	0 P

Analysis of Variance	DF	F-incr	F-adj	StdErrDiff
12 mu	1	11350.16	0.33	
5 year	5	46.07	23.51	
6 sex	1	9.66	6.19	0.1623
7 agewean	1	63.84	58.59	
8 damage	1	30.44	98.43	
10 damagsq	1	78.41	83.23	
4 sirebrd	1	6.64	5.54	0.1755
2 dambrd	1	2.91	2.91	0.2664

	Solution	Standard Error	T-value	T-prev
2 dambrd				
	2	-0.454301	0.266447	-1.71
4 sirebrd				
	4	-0.413039	0.175532	-2.35
10 damagsq				
	5	-0.289973	0.317842E-01	-9.12
8 damage				
	6	2.92231	0.294547	9.92
7 agewean				
	7	0.659286E-01	0.861291E-02	7.65
6 sex				
	9	-0.403814	0.162311	-2.49
5 year				
	101	-1.57091	0.267785	-5.87
	102	-1.07664	0.264296	-4.07
	103	-3.00250	0.344567	-8.71
	104	-3.28832	0.345212	-9.53
	105	-2.45008	0.394630	-6.21
12 mu				
	106	0.589612	1.03227	0.57
3 sire	74 effects fitted			
1 dam	358 effects fitted			

The output above gives the same parameter estimates as the one given earlier by Genstat. The Wald statistics are now replaced by F-values and the ‘Tables of effects’ with their respective standard errors appear under ‘solution’.

Note that this output has two forms of F-values, namely ‘F-incr’ and ‘F-adj’. The former stands for F incremental – in other words each F value is that corrected for the variables already included, whereas ‘F-adj’ stands for F adjusted – i.e. here F value is calculated adjusted for the other terms in the model regardless of the order in which they are fitted.

Heritability estimates and their standard errors can be additionally calculated with ASREML to give:

direct heritability 0.054 ± 0.071
maternal heritability 0.294 ± 0.050

The formulae used for this example are:

direct heritability = $4\sigma_{\text{sire}}^2 / (\sigma_{\text{sire}}^2 + \sigma_{\text{dam}}^2 + \sigma_{\text{error}}^2)$
where σ^2 is the variance component associated with the subscript
maternal heritability = $\sigma_{\text{dam}}^2 / (\sigma_{\text{sire}}^2 + \sigma_{\text{dam}}^2 + \sigma_{\text{error}}^2)$

These heritability estimates demonstrate the strong maternal influence on weaning weight and the weak genetic inheritance from the parent.

Study questions

1. Discuss the advantages and disadvantages of mixed models analysis compared with least squares analysis of variance.
2. Before REML procedures were available, statistical analysis was often undertaken in two stages -first least squares analysis to estimate fixed effects, then least squares analysis to estimate random effects corrected for fixed effects. (It was difficult to mix the two). What

do you think are the advantages of REML over this approach? Can you think of any occasion when correcting data for fixed effects might still be useful?

3. In the following output year of birth and year x breed are considered as random terms in the model. What inferences can be made about differences between breeds? Discuss whether you think it appropriate to make the assumptions that years can be represented as a random sample.

REML Variance Components Analysis

Response Variate : WEANWT

Fixed model: Constant+SEX+AGEWEAN+DL+DQ+SIRE_BRD+DAM_BRD

Random model: SIRE_ID+DAM_ID+YEAR+YEAR.BREED

Number of units: 700

Estimated Variance Components

Random term	Component	S.e.
SIRE ID	0.065	0.089
DAM ID	1.452	0.283
YEAR	1.462	0.963
YEAR.BREED	0.021	0.066
units	3.421	0.267

Approximate stratum variances

	Effective	d.f.
SIRE ID	4.739	57.60
DAM ID	6.480	297.74
YEAR	224.485	4.81
YEAR.BREED	4.372	4.18
units	3.421	328.67

Matrix of coefficients of components for each stratum

SIRE ID	10.34	0.42	0.00	1.06	1.00
DAM ID	0.00	2.10	0.00	0.30	1.00
YEAR	0.00	0.00	150.48	54.42	1.00
YEAR.BREED	0.00	0.00	0.00	45.99	1.00
units	0.00	0.00	0.00	0.00	1.00

Deviance: -2*Log-Likelihood

Deviance d.f.

1825.94 689

Note: deviance omit constants which depend on fixed model fitted.

Wald tests for fixed effects

Fixed term	Wald statistic	d.f.
SEX	9.7	1
AGEWEAN	65.8	1
DL	28.0	1
DQ	78.2	1
SIRE_BRD	5.5	1
DAM_BRD	3.8	1

* All Wald statistics are calculated ignoring terms fitted later in the model

Table of effects for SIRE_BRD

SIRE_BRD	D	R
	0.0000	-0.3981
Standard error of differences:	0.1892	

Table of effects for DAM_BRD

DAM_BRD	D	R
	0.0000	-0.5296
Standard error of differences:	0.273	

- The analysis of weaning weight demonstrates a small sire variance component. Describe what additional model you might run and how you would use the two outputs to evaluate the significance of the sire variance component
- Express the results of the mixed model analysis described in this module in a way that you might present in a scientific paper. Include a summary of the statistical analysis, a table of results and a short text to describe the results.
- The REML analysis using the ASREML program for weight at one year is given below. The analysis excludes missing values probably resulting from lambs that died before the age of one year. Interpret it. Does this provide a more suitable trait for estimating direct heritability of growth? Explain why.

ASREML Diani data						
Source	Model	terms	Gamma	Component	Comp/SE	% C
Sire	5338	74	0.477068E-01	0.139349	1.28	0 P
Dam	12682	22	0.517507	1.51161	5.05	0 P
Variance	598	586	1.00000	2.92094	11.24	0 P

Analysis of Variance	DF	F-incr	F-adj	StdErrDiff
12 mu	1	10977.64	10.93	
5 year	5	56.66	31.72	
6 sex	1	27.02	21.45	0.1668
7 agewean	1	53.20	51.35	
9 damage	1	21.55	57.98	
10 damagsq	1	44.54	48.42	
4 sirebrd	1	7.78	6.83	0.1935
2 dambrd	1	2.45	2.45	0.2901

	Solution	Standard Error	T-value	T-prev
2 dambrd				
	2 -0.454354	0.290130	-1.57	
4 sirebrd				
	4 -0.505866	0.193494	-2.61	
10 damagsq				
	5 -0.237768	0.341697E-01	-6.96	
9 damage				
	6 2.39436	0.314442	7.61	
7 agewean				
	7 0.625601E-01	0.873008E-02	7.17	
6 sex				
	9 -0.772542	0.166795	-4.63	
5 year				
	101 -2.15099	0.266150	-8.08	
	102 -1.89963	0.286131	-6.64	0.91
	103 -4.05338	0.353991	-11.45	-7.02
	104 -3.38650	0.380307	-8.90	2.09
	105 -4.11357	0.423752	-9.71	-1.98
12 mu				
	106 3.58347	1.08413	3.31	
3 sire	74 effects fitted			
1 dam	322 effects fitted			

This yields the following heritability estimates:

```

4 phen 1 4.572 0.2869
5 direct 1 0.5574 0.4364
herit = direct 5/phen 1 4 = 0.1219 0.0945
m2 = dam 2/phen 1 4 = 0.3306 0.0553

```