Chapter 8
Hypothesis testing in mixed models

Julius van der Werf

Variance of predictors and prediction errors

A prediction from a mixed model uses a combination of estimates of fixed effects and predictions of random effects. For example, we can predict the performance of a certain daughter of a bull in a certain herd at a certain age.

The model is: $y = \mu + b + u + e$.

The *predictand* is K'b + M'u
The *predictor* is a linear function of y, i.e. L'y (practically, a linear combination of the estimated parameters, which in themselves are linear functions of the data)

The Prediction error is the difference between the predictor and the predictand, i.e. K'b + M'u – L'y. If the expectation of this is zero, then the prediction is unbiased.

The prediction error variance:
$V(b - \hat{b}) = V(\hat{b}) = (X'V^{-1}X)^{-}$ the square root of this value is the SE

$V(u - \hat{u}) = V(\hat{u}) + V(u) - 2Cov(u, \hat{u})$

$= V(u) - V(\hat{u})$ as $Cov(u, \hat{u}) = V(\hat{u})$

$= G - V(\hat{u})$

Further: $Cov(\hat{b}, u - \hat{u}) = 0$, and $Cov(\hat{b}, \hat{u}) = 0$

Note that with more information, the PEV decreases. However, $V(\hat{b})$ decreases with more information whereas $V(\hat{u})$ increases. When information → 8, then $V(\hat{b})$ → 0 whereas $V(\hat{u})$ → G

These PEVs can best be obtained from the mixed model equations. The solutions to the MME can be written as

$$\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} C_{XX} & C_{XZ} \\ C_{ZX} & C_{ZZ} \end{bmatrix} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Where the matrix is the generalized inverse of the coefficient matrix of the MME, i.e. C_{XX} is the 'fixed effects part' of the inverse, and NOT the inverse of $XR^{-1}X$

Now, $\text{Var}(\hat{b}) = C_{XX}$

$V(\hat{u}) = G - C_{ZZ}$

And $V(u - \hat{u}) = C_{ZZ}$

PEV's of estimated breeding values

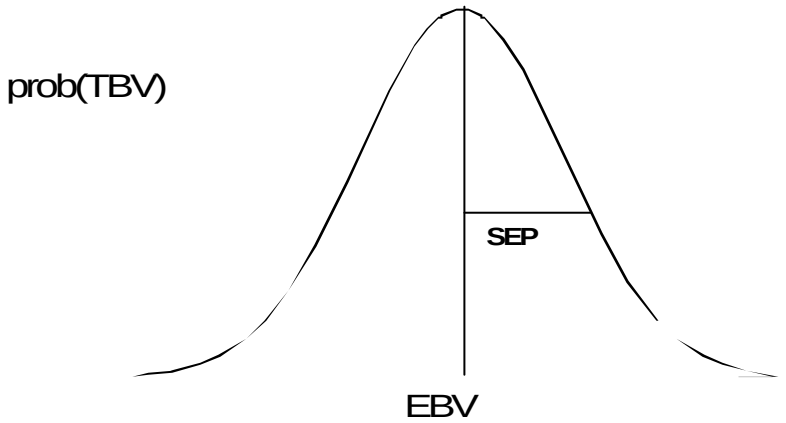
In a BLUP model we can have animals' additive genetic effects as random effects. Now the $V(\hat{u})$ = the variance of the EBV's. From quantitative genetic theory we know that var(EBV) = $r_{IH}^2 V_A$, where r_{IH} is the accuracy of the EBV and V_A is the additive genetic variance. From the BLUP model we can first obtain the diagonal element of the inverse C_{ZZ} (sometimes we approximate this value as the MME coefficient matrix is often not inverted), for animal I this is C^{ii}.

The Prediction Error Variance of the EBV: $PEV = C^{ii}$

This is also equal to $(1 - r_{IH}^2) V_A$

Note that var(EBV) + PEV add to VA

(Note that MME often have multiplied out R^{-1}, i.e. we use Z'Z rather than $Z'R^{-1}Z$ etc. In that case the Prediction Error Variance of the EBV: $PEV = C^{ii}\sigma_e^2$ **)**



The Standard Error of Prediction (SEP) is $\sqrt{}$(PEV). ASREML gives SEP values behind the solution of random effects (*.sln file)

The reliability of breeding values can be calculated as r_{IH}^2= 1-(PEV/VA) and the accuracy is the square root of this.

Think again about the extreme cases:

- when there is no information, and accuracy is 0: all EBV's will then be 0 and the variance of the prediction error PEV = VA.
- when there is full information, the EBV will be equal to the true BV and the variance of the prediction error PEV = 0.

The prediction error of an EBV is important since it gives us a clue of how far the true breeding value could be off the EBV. This is important for example to answer questions like: how much could an EBV still change if we obtain more information on the animal. Changes in EBV's are not good for the industry's confidence in the genetic evaluation system. However, we have to realise that an EBV is never exact, unless the accuracy is 100%. We expect the true breeding value to be the same as the EBV, but there is a certain probability that it will be a bit different. The probability distribution of the true EBV, given an EBV looks like in the figure.

Hypothesis Testing in Mixed Models

Hypothesis testing in the case of mixed models with unbalanced data is not well understood. Many analyse the fixed effects only and ignore random effects. Other treat random effects (e.g. sires) as fixed. In hypothesis testing, expectations are derived assuming the true model. However, the variance components needed in a mixed model are estimates, and therefore strictly solutions for fixed effects (combinations) are not BLUE.

<u>If G and R are known</u>,
then V is also known and estimates of b are BLUE and the hypothesis test is as described before. To test H'b-c = 0

We used the test statistic $F = \dfrac{s / r(H')}{SSE/(N - r(X))}$

where $s = (H'\beta\text{-}c)'(H'CH)^{-1}(H'\beta\text{-}c)$ and $C = (X'V^{-1}X)^{-1}$

This test is exact and best, given that G and R and known, or known to proportion.

<u>When G and R are not known</u>,
there is no best test and BLUE of b is not possible. If estimates of G and R are used, then hypothesis testing is only approximate. The possibilities are:

a) Estimation by computing as though random effects were fixed:

$$\begin{bmatrix} b \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

For this case, <u>if K'b is estimable</u>, $K'\beta$ is an *unbiased* estimator of k'b., however, it is not a minimum variance estimator of b and it does not have maximum power.
An exact test is given by: $Q/f\hat{s}_e^2 \sim F[f, N\text{-}r(X,Z)]$
for $Q = (K'\beta - m)'(K'C_{11}K)^{-1}(K'\beta m)$ and $\hat{s}_e^2$ is the estimated residual variance from the model.

If K'b is not estimable, no exact test exists. The estimate k'b depends on the choice of m'u, and the F-test can be inflated because the denominator (containing only residual variance) is too small. The degree of bias in F depends on the ratio of variance components.

b) Estimation ignoring all random effects

$\beta = (X'X)^{-}X'y$

If K'b is estimable, K'β is an *unbiased* estimator of k'b., however, it is not a minimum variance estimator of b and it does not have maximum power. No exact test is possible. F-tests are often inflated if K'b is not estimable. An approximate test can improve the properties of the test but few statistical packages would accommodate this.

c) Estimation by computing with estimates of the variances of the random effects

$$\begin{bmatrix} b \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \hat{G}^{-1} \end{bmatrix}^{-} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

This approach gives unbiased estimates and often with smaller sampling error than when treating random effects as fixed. The F test is approximate, with better properties if the estimated value for G approximates the true value. Effectively the residual variance is corrected for random effects. The denominator of the F-test contains residual variance and a term for the variance components. This F-test is more precise and the denominator is not inflated as in a) or b). If no good estimates of variance components exist, however, it might be safer to follow approach 1)

d) Simultaneous estimation of fixed effects and variance components for random effects, e.g. using REML. This gives an approximated F-test as in c). This is often the most sensible approach, given that there is a reasonable amount of data to estimate variance components. ASREML will provide the most appropriate F statistic for this case.

Exercises:

1. Using mixed models to test treatment effects

Many experiments in animal science use repeated measures on animals. Often, covariances among repeated measuremens on the same animal are not taken into account when the interest is only in estimating the treatment effect (e.g. in a nutrition trial). We will look at the possible problems with that approach.

Consider the following data set with two treatments measured on 4 cows (2 for each treatment) with 5 repeated measurements per cow (i.e. 10 measurements per treatment.

Treatment	Cow	
I	A	451; 456; 462; 449; 455;
	B	472; 469; 476; 467; 462;
II	C	481; 475; 482; 489; 483;
	D	510; 502; 499; 507; 501;

Test the treatment effect, with and without cow fitted.

2. Principles of mixed models and BLUP

The principle of estimation of breeding value is based on using

- Phenotypic observations as deviations to expected means (e.g. contemporary group mean)
- Weighing those deviations with a regression coefficient.

In selection index, the expected means are assumed known, and the only task of breeding value estimation is to find the appropriate weights for deviations (i.e. deviations of the observed records from those means).

In BLUP, these expected means (of a contemporary group) have to be estimated from the data.
A linear (mixed) model is used for this purpose.

> A mixed model is a linear model for fixed and random effects. Breeding values are random effects (they have variation). Random effects are estimated differently from fixed effects. Fixed effects are basically estimated as observed means (possibly corrected from some other fixed effects), whereas random effects are regressed toward a certain mean. If a herd mean is +1 above the breed average, we believe it is all herd effect and the herd is expected to be really one unit better then others. If an animal is +1 above the herd mean, we believe only part of this is due to the animals' genes, the other part is due to the environmental (or error) effect. The distinction between fixed and random effects is hard and would need a more theoretical statistical coverage

This practical will show by a simple example that both methods use the same weights for amalgamating the different pieces of information to estimate and EBV.

Given is data on 6 animals with the following information

animal	sire	herd	observed performance
1	0	1	35
2	0	2	50
3	2	1	42
4	2	1	38
5	1	2	47
6	1	2	43

a) Calculate the EBV for animal 1, using selection index (you can use STSELIND for this purpose to derive the weights, or, if you feel challenged, derive them yourself)

b) Calculate the EBV of all animals using a random model.

Work this out as follows (you can work in excel)

- Let the observed data be in **y** (a vector of 6 x 1)
- Express them as deviations from the mean: this is a new vector called $\mathbf{y^d}$.
  In a random model we only estimate random effects, but we estimate them all jointly.
  The random model looks like

$$\mathbf{y^d} = y - Xb$$
$$= Zu + e$$

- Now work out how the matrix Z looks like (it links the data up to the animals)
- Work out the additive genetic relationships between all animals, call this matrix **A**

- The breeding values can be estimated from $[Z'Z + lA^{-1}][\hat{u}] = Z'y^d$

so that $\hat{u} = [Z'Z + lA^{-1}]^{-} Z'y^d$

work in steps:

- determine the matrix $Z'Z + \lambda.A^{-1}$
- invert this matrix
- multiply with the vector $Z'y^d$

This gives you the EBV for each animal, they are in the vector *u-hat*

- Now look also at the elements in the inverse of $[Z'Z + \lambda.A^{-1}]$ and compare these with the index weights in 1)

c) Calculate the EBV of all animals using a mixed model.

d) Now swap animal 4 and 5 over the herds and compare the EBV's and their accuracies for the two cases.