

Chapter 12

Variance Components

The prediction of breeding values using a BLUP animal model required the **variance components** σ_e^2 for the residual variance and σ_u^2 for the genetic additive variance to be known. For the sire model, σ_u^2 is replaced by the sire variance component σ_s^2 . In real world livestock breeding evaluations, these variance components are not known and hence must be estimated from the data. The data analysis procedure that estimates the variance components from data is called **variance components estimation**.

12.1 Sire Model

The sire model is used to motivate the introduction of the topic of variance components estimation. The sire model is given by

$$y = X\beta + Z_s s + e \quad (12.1)$$

with $\text{var}(e) = R$, $\text{var}(s) = A_s \sigma_s^2$ and $\text{var}(y) = Z_s A_s Z_s^T \sigma_s^2 + R$. The matrix A_s is the numerator relationship for sires, the sire variance component σ_s^2 corresponds to $0.25 * \sigma_u^2$ and R can often be simplified to $R = I * \sigma_e^2$. The interest in this chapter is how to estimate σ_s^2 and σ_e^2 .

In the simple case the vector β is reduced to just one scalar fixed effects parameter. This reduced X to a matrix with one column with all elements equal to 1. Assuming that we have q unrelated sires the relationship matrix A_s for the sires corresponds to the identity matrix I .

12.2 Analysis Of Variance (Anova)

As a first approach we can use an analysis of variance by fitting

1. a model with an overall effect $\beta = \mu$ and
2. a model with sire effects.

These two models give an analysis of variance of the following structure

Source	Degrees of Freedom (<i>df</i>)	Sums of Squares (<i>SSQ</i>)
Overall (μ)	$Rank(X) = 1$	$y^T X (X^T X)^{-1} X^T y = F$
Sires (s)	$Rank(Z_s) - Rank(X) = q - 1$	$y^T Z_s (Z_s^T Z_s)^{-1} Z_s^T y - y^T X (X^T X)^{-1} X^T y = S$
Residual (e)	$n - Rank(Z_s) = n - q$	$y^T y - y^T Z_s (Z_s^T Z_s)^{-1} Z_s^T y = T$
Total	n	$y^T y$

The sums of squares (*SSQ*) can also be expanded into sums of scalar quantities which might be easier to understand. For our sire model we get

$$F = y^T X (X^T X)^{-1} X^T y = \frac{1}{n} \left[\sum_{i=1}^n y_i \right]^2$$

where n corresponds to the number of observations in the dataset.

$$S = y^T Z_s (Z_s^T Z_s)^{-1} Z_s^T y - y^T X (X^T X)^{-1} X^T y = \sum_{i=1}^q \frac{1}{n_i} \left[\sum_{j=1}^{n_i} y_{ij} \right]^2 - F$$

where n_i corresponds to the number of observations for sire i .

$$T = y^T y - y^T Z_s (Z_s^T Z_s)^{-1} Z_s^T y = \sum_{i=1}^n y_i^2 - S - F$$

In principle effects β and s are treated as fixed effects in the above anova. If estimates of σ_e^2 and σ_s^2 are required the observed sums of squares S and T can be equated to their expected values $E(T) = (n - q)\sigma_e^2$ and $E(S) = (q - 1)\sigma_e^2 + tr(Z_s M Z_s)\sigma_s^2$ where $M = I - X(X^T X)^{-1} X^T$ and $tr(M)$ stands for the trace of matrix M which corresponds to the sum of the diagonal elements of matrix M .

12.3 Numerical Example

We want to show the estimation of variance components with a very small data set. The data that will be used is shown in the table below. The observations consist of pre-weaning weight gains of beef cattle.

Table 12.1: Small Example Dataset for Variance Components Estimation Using a Sire Model

Animal	Sire	WWG
4	2	2.9
5	1	4.0
6	3	3.5
7	2	3.5

The model used is a simplified sire model where all the fixed effect are captured by a common mean μ . Then there is the sire effect s as a random effect and the random residual effect. Hence for any given observation y_{ij} for animal i of sire j , we can write

$$y_{ij} = \mu + s_j + e_i$$

with μ the common mean, s_j the random effect of sire j ($j = 1, 2, 3$) and e_i corresponds to the random residual of observation i ($i = 1, \dots, 4$). In matrix notation this model was already given in (12.1). The design matrix X is a matrix with one column and with elements all equal to 1. The design matrix Z_s links observations to sire effects.

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, Z_s = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

An analysis of variance can be constructed as

Source	Degrees of Freedom (df)	Sums of Squares (SSQ)
Overall (μ)	$Rank(X) = 1$	$F = 48.3025$
Sires (s)	$Rank(Z_s) - Rank(X) = q - 1$	$S = 0.4275$
Residual (e)	$n - Rank(Z_s) = n - q$	$T = 0.18$

With

$$M = \begin{bmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix} \text{ and } Z_s^T M Z_s = \begin{bmatrix} 0.75 & -0.50 & -0.25 \\ -0.50 & 1.00 & -0.50 \\ -0.25 & -0.50 & 0.75 \end{bmatrix}$$

we get the following estimates

$$\hat{\sigma}_e^2 = T = 0.18$$

$$\hat{\sigma}_s^2 = \frac{S - (q - 1)\hat{\sigma}_e^2}{\text{tr}(Z_s^T M Z_s)} = \frac{0.4275 - 2 * 0.18}{2.5} = 0.027$$

The same computations based on an anova can be done in R very easily. Assume that our dataset is in a dataframe which is called `tbl_num_ex_chp12_aov`. We are doing the anova using the function `aov()` to get the sums of squares.

```
aov_num_ex_chp12 <- aov(formula = WWG ~ Sire, data = tbl_num_ex_chp12_aov)
summary(aov_num_ex_chp12)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Sire       2  0.4275  0.2137   1.187  0.544
## Residuals  1  0.1800  0.1800
```

The results from above are obtained for $\hat{\sigma}_e^2 = 0.18$ as the value under the column `Mean Sq` in the row `Residuals`. Because in our computations above, we have considered the estimation of the overall effect which is not done in the function `aov()` in R.

12.4 Negative Estimates with Anova

One of the problems that frequently occurs when using anova to estimate variance components is that some estimates might be negative. Negative estimates are outside of the permissible range for the parameter and hence are not valid estimates. As a consequence of that alternative methods have been proposed to estimate variance components.

12.5 Likelihood-Based Approaches

The maximum likelihood (ML) approach was developed and popularized by R. A. Fisher. ML is a general approach for parameter estimation and is not only used for estimating variance components. Let us assume that our observed traits are continuous and real-valued quantities. In ML we assume that these quantities follow a certain density. This density is a function of the observed values and of unknown parameters that we want to estimate.

12.5.1 Density of Observations

Given a vector y of observations. As already mentioned, the vector y follows a certain density. As an example such a density might be a multivariate normal distribution. For a given vector y of length n , the underlying n -dimensional multivariate normal distribution has the following form

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp \left\{ -\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) \right\}$$

with μ expected value of y
 Σ variance-covariance matrix of y
 $\det()$ determinant

12.5.2 Likelihood Function

As already mentioned the density is a function of the observed data y and of some unknown parameters. For the multivariate normal distribution these parameters are μ and Σ . Before observing any data, we can interpret the density $f(y|\mu, \Sigma)$ as a function of y for some fixed values of μ and Σ . But once the data has been observed, y is fixed and the parameters μ and Σ are unknown and must be estimated from the data. For the task of parameter estimation, it makes more sense to view $f(y|\mu, \Sigma)$ as a function of μ and Σ . We can write this function a little different

$$L(\mu, \Sigma) = f(y|\mu, \Sigma)$$

The function $L(\mu, \Sigma)$ is called the **Likelihood** function.

12.5.3 Maximum Likelihood

For a given dataset we choose an appropriate density which is suitable for our observations. As already mentioned, due to the Central Limit Theorem, the normal distribution is often used as a density for observations. Once, we have chosen the density, it contains unknown parameters which we have to estimate from the data. Loosely speaking, our goal is to determine the parameters such that the observed data is modeled as good as possible. This requirement is translated into a mathematical framework by the maximization of the likelihood. Hence for a given dataset our parameter estimates are determined such that the likelihood is maximized. For our multi-variate normal distribution, this can be transformed into the following equations

$$\hat{\mu} = \operatorname{argmax}_{\mu} L(\mu, \Sigma)$$

and

$$\hat{\Sigma} = \operatorname{argmax}_{\Sigma} L(\mu, \Sigma)$$

12.6 Summary

The topic of variance component estimation is a huge area. We have just covered two possible approaches to get estimates of variance components. There are many more of them. The coverage of these methods is outside of the scope of this course.