

# 20. REML Estimation of Variance Components

Consider the General Linear Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

and  $\Sigma$  is an  $n \times n$  positive definite variance matrix that depends on unknown parameters that are organized in a vector  $\gamma$ .

- In the previous set of slides, we considered maximum likelihood (ML) estimation of the parameter vectors  $\beta$  and  $\gamma$ .
- We saw by example that the MLE of the variance component vector  $\gamma$  can be biased.

## Example of MLE Bias

For the case of  $\Sigma = \sigma^2 \mathbf{I}$ , where  $\gamma = \sigma^2$ , the MLE of  $\sigma^2$  is

$$\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$

with expectation

$$\frac{n-r}{n} \sigma^2.$$

This MLE for  $\sigma^2$  is often criticized for “failing to account for the loss of degrees of freedom needed to estimate  $\beta$ .”

$$\begin{aligned} \text{E} \left[ \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n} \right] &= \frac{n-r}{n} \sigma^2 \\ &< \sigma^2 \\ &= \text{E} \left[ \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{n} \right]. \end{aligned}$$

*known*

## A Familiar Special Case

$$y_1, \dots, y_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\mathbb{E} \left[ \frac{\sum_{i=1}^n (y_i - \mu)^2}{n} \right] = \sigma^2 \text{ but}$$

$$\mathbb{E} \left[ \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} \right] = \frac{n-1}{n} \sigma^2 < \sigma^2.$$

- REML is an approach that produces unbiased estimators for these special cases and produces less biased estimates than ML in general.
- Depending on whom you ask, REML stands for REidual Maximum Likelihood or REsstricted Maximum Likelihood.

# The REML Method

$$= r$$

- ① Find  $n - \underline{\text{rank}}(\mathbf{X}) = n - r$  linearly independent vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{n-r}$  such that  $\mathbf{a}_i^\top \mathbf{X} = \mathbf{0}^\top$  for all  $i = 1, \dots, n - r$ .
- ② Find the maximum likelihood estimate of  $\gamma$  using  $w_1 \equiv \mathbf{a}_1^\top \mathbf{y}, \dots, w_{n-r} \equiv \mathbf{a}_{n-r}^\top \mathbf{y}$  as data. — *create a "new" data set*

$$\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_{n-r}] \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_{n-r} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{y} \\ \vdots \\ \mathbf{a}_{n-r}^\top \mathbf{y} \end{bmatrix} = \mathbf{A}^\top \mathbf{y}$$

*new data*

- If  $a^\top X = 0^\top$ ,  $\boxed{a^\top y}$  is known as an error contrast.
- Thus,  $w_1, \dots, w_{n-r}$  comprise a set of  $n - r$  error contrasts.
- Because

$$\underbrace{(I - P_X)X}_{A^\top} = X - P_X X = X - X = 0,$$

*estimated residuals*

the elements of

$$(I - P_X)\hat{y} = y - P_X y = y - \hat{y}$$

are each error contrasts.

- Because  $\text{rank}(\mathbf{I} - \mathbf{P}_X) = n - r$ , there exists a set of  $n - r$  linearly independent rows of  $\mathbf{I} - \mathbf{P}_X$  that can be used in step 1 of the REML method to get  $a_1, \dots, a_{n-r}$ .  $\leftarrow$  reflects the # of elements in our  $\beta$
- If we do use a subset of rows of  $\mathbf{I} - \mathbf{P}_X$  to get  $a_1, \dots, a_{n-r}$ ; the error contrasts

$$w_1 = \mathbf{a}_1^\top \mathbf{y}, \dots, w_{n-r} = \mathbf{a}_{n-r}^\top \mathbf{y}$$

will be a subset of the elements of the residual vector

$$(\mathbf{I} - \mathbf{P}_X)\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}}.$$

- This is why it makes sense to call the procedure Residual Maximum Likelihood.

Distribution of  $\hat{\omega} = A^T y$

Note that

$$\begin{aligned} w &= A^T y \\ &= A^T(X\beta + \epsilon) \\ &= \boxed{A^T X}\beta + A^T \epsilon \\ &= \mathbf{0} + A^T \epsilon \\ &= A^T \epsilon \end{aligned}$$

$w$  is a random vector depending on some constant  $A$

Thus,

$$w = A^T \epsilon \sim \mathcal{N}(A^T \mathbf{0}, A^T \Sigma A) \stackrel{d}{=} \mathcal{N}(\mathbf{0}, A^T \Sigma A), \text{ and the}$$

and the distribution of  $w$  depends on  $\gamma$  but not  $\beta$ . residual

vector  $\epsilon$ :  
from slide 1:  $\epsilon \sim N(0, \Sigma)$

key idea:  $\omega$  contains all information contained in our data  $\underline{\Sigma}$  has a mean = 0  
 $\Rightarrow$  no longer need to estimate the mean

The log likelihood function in this case is      Structure when using MLE of

$$\ell(\gamma|\mathbf{w}) = -\frac{1}{2} \log |\mathbf{A}^\top \Sigma \mathbf{A}| - \frac{1}{2} \mathbf{w}^\top (\mathbf{A}^\top \Sigma \mathbf{A})^{-1} \mathbf{w} - \frac{n-r}{2} \log(2\pi).$$

An MLE for  $\gamma$ , say  $\hat{\gamma}$ , can be found in the general case using numerical methods to obtain the REML estimate of  $\gamma$ .

In 611.0 we take the time to prove that every set of  $n - r$  linearly independent error contrasts yields the same REML estimator of  $\gamma$ .

As an example, consider the special case where

$$y_1, \dots, y_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2).$$

Then  $X = 1$ ,  $\beta = \mu$ , and  $\Sigma = \sigma^2 I$ .

$$\text{rank}(X) = 1 \implies n-1 \text{ error contrasts}$$

$$A^T = (\mathbb{I} - P_X) \quad \text{base on } X = 1$$

It follows that

$$\begin{aligned} a_1^T &= (1, -1, 0, 0, \dots, 0) \\ a_2^T &= (0, 1, -1, 0, \dots, 0) - y_2 - y_3 \\ &\vdots \\ a_{n-1}^T &= (0, 0, \dots, 0, 1, -1) - \dots - y_{n-1} - y_n \end{aligned}$$

and

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \begin{aligned} a_1^T y &= (1, -1, \dots, 0) \\ a_1^T y &= y_1 - y_2 \end{aligned}$$

representing  $y_n$

$y_1 - y_n$

$$\mathbf{b}_1^\top = (1, 0, 0, \dots, 0, -1)$$

$$\mathbf{b}_2^\top = (0, 1, 0, \dots, 0, -1)$$

$\vdots$

$$\mathbf{b}_{n-1}^\top = (0, 0, \dots, 0, 1, -1) - \dots - y_{n-1} - y_n$$

are each a set of  $n - r = n - 1$  linear independent vectors that can be used to form error contrasts.

Either

$$w = \begin{bmatrix} a_1^\top \mathbf{y} \\ a_2^\top \mathbf{y} \\ \vdots \\ a_{n-1}^\top \mathbf{y} \end{bmatrix} = \begin{bmatrix} y_1 - y_2 \\ y_2 - y_3 \\ \vdots \\ y_{n-1} - y_n \end{bmatrix} \text{ or } v = \begin{bmatrix} b_1^\top \mathbf{y} \\ b_2^\top \mathbf{y} \\ \vdots \\ b_{n-1}^\top \mathbf{y} \end{bmatrix} = \begin{bmatrix} y_1 - y_n \\ y_2 - y_n \\ \vdots \\ y_{n-1} - y_n \end{bmatrix}$$

could be used to obtain the same REML estimator of  $\sigma^2$ .

REML = MLE but based on an  
"adjusted" data set based on

For the normal theory Gauss-Markov linear model, the original  
data

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

the REML estimator of  $\sigma^2$  is

$$\hat{\sigma}^2 = \frac{\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_X) \mathbf{y}}{n - r},$$

because our

newly  
created data

the unbiased estimator that we used previously.

Set consists

of  $n-1$  observ.  
instead of  $n$ .

For linear mixed effects models, the REML estimators of variance components produce the same estimates as the unbiased ANOVA-based estimators formed by taking appropriate linear combinations of mean squares when the latter are positive and data are balanced.

REML = ANOVA based  
estimators

In any case, once a REML estimate of  $\gamma$  (and thus  $\Sigma$ ) has been obtained, the BLUE of an estimable  $C\beta$  can be approximated by

$$C\hat{\beta}_{\hat{\Sigma}} = C(\mathbf{X}^\top \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Sigma}^{-1} \mathbf{y},$$

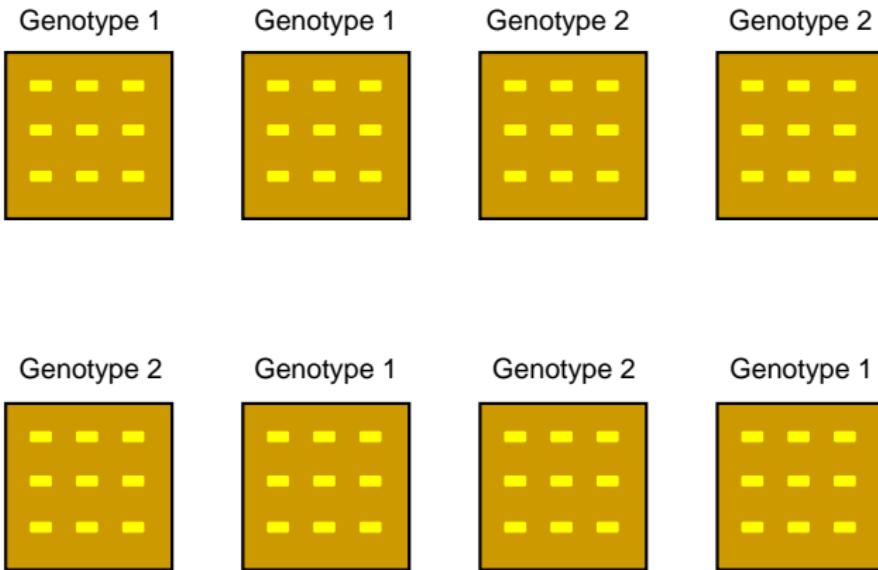
where  $\hat{\Sigma}$  is  $\Sigma$  with  $\hat{\gamma}$  (the REML estimate of  $\gamma$ ) in place of  $\gamma$ .

approximate

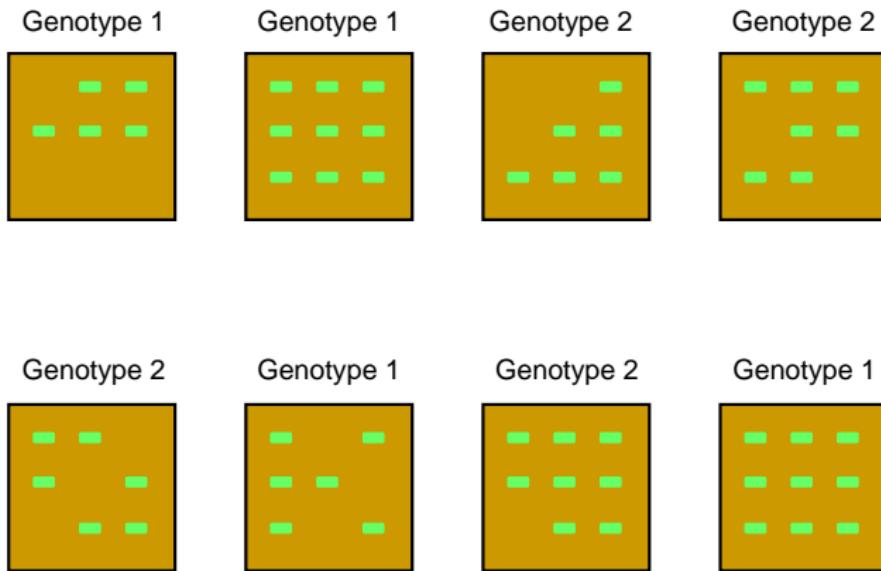
## An Example

Researchers were interested in comparing the dry weight of maize seedlings from two different genotypes. For each genotype, nine seeds were planted in each of four trays. The eight trays in total were randomly positioned in a growth chamber. Three weeks after the emergence of the first seedling, emerged seedlings were harvested from each tray and individually weighed after drying to obtain one dry weight for each seedling. Although nine seeds were planted in each tray, fewer than nine seedlings emerged in many of the trays.

# Planted Seeds



# Emerged Seedlings

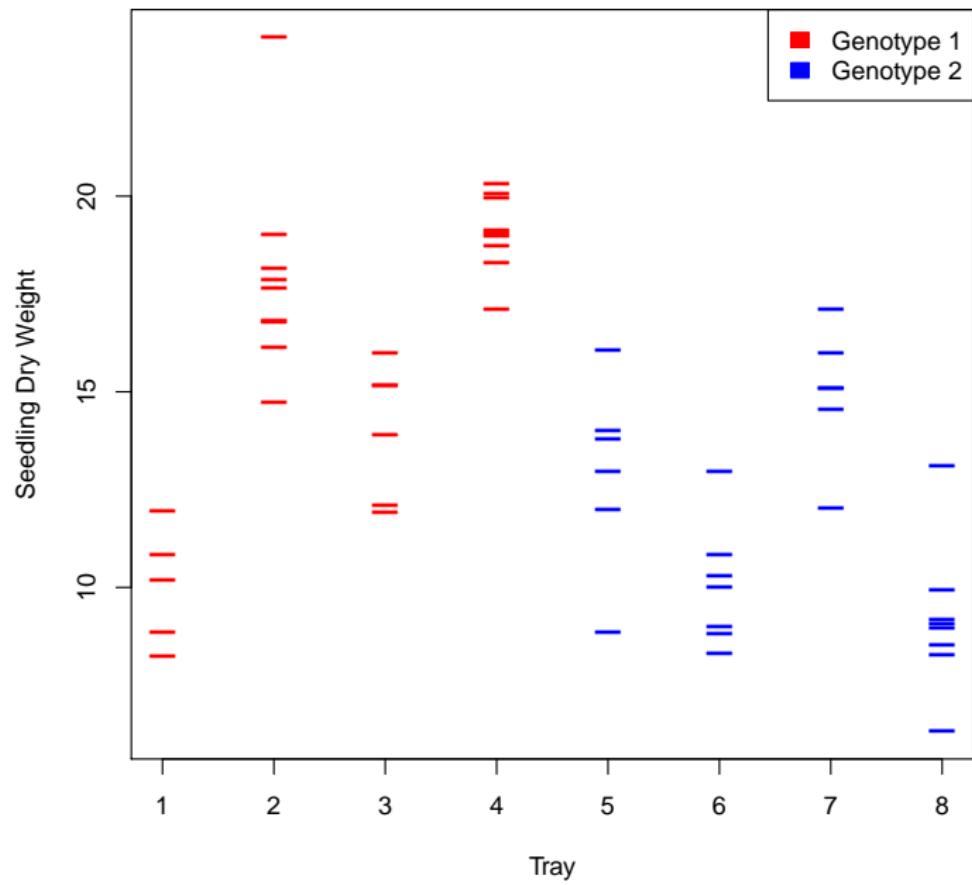


```
> d=read.delim(  
+ "https://dnett.github.io/S510/SeedlingDryWeight2.txt")  
> d  
  
Genotype Tray Seedling SeedlingWeight  
1      1     1       1        8  
2      1     1       2        9  
3      1     1       3       11  
4      1     1       4       12  
5      1     1       5       10  
6      1     2       1       17  
7      1     2       2       17  
8      1     2       3       16  
9      1     2       4       15  
10     1     2       5       19  
11     1     2       6       18  
12     1     2       7       18  
13     1     2       8       18  
14     1     2       9       24  
15     1     3       1       12
```

16	1	3	2	12
17	1	3	3	16
18	1	3	4	15
19	1	3	5	15
20	1	3	6	14
21	1	4	1	17
22	1	4	2	20
23	1	4	3	20
24	1	4	4	19
25	1	4	5	19
26	1	4	6	18
27	1	4	7	20
28	1	4	8	19
29	1	4	9	19
30	2	5	1	9
31	2	5	2	12
32	2	5	3	13
33	2	5	4	16
34	2	5	5	14

35	2	5	6	14
36	2	6	1	10
37	2	6	2	10
38	2	6	3	9
39	2	6	4	8
40	2	6	5	13
41	2	6	6	9
42	2	6	7	11
43	2	7	1	12
44	2	7	2	16
45	2	7	3	17
46	2	7	4	15
47	2	7	5	15
48	2	7	6	15
49	2	8	1	9
50	2	8	2	6
51	2	8	3	8
52	2	8	4	8
53	2	8	5	13
54	2	8	6	9
55	2	8	7	9
56	2	8	8	10

```
> plot(d[,2],d[,4]+rnorm(56,0,.2),
+       xlab="Tray",ylab="Seedling Dry Weight",
+       col=2*d[,1],pch="-",cex=2)
> legend("topright",c("Genotype 1","Genotype 2"),
+        fill=c(2,4),border=c(2,4))
```



## A Model for the Seedling Dry Weights

Let  $y_{ijk}$  be the dry weight of the  $k$ th seedling in the  $j$ th tray for genotype  $i$ .

Suppose

$$y_{ijk} = \underline{\mu_i} + \underline{t_{ij}} + e_{ijk},$$

where  $\mu_1$  and  $\mu_2$  are unknown constants,

$$\underline{t_{ij}} \sim \mathcal{N}(0, \sigma_t^2), \quad \underline{e_{ijk}} \sim \mathcal{N}(0, \sigma_e^2),$$

and all random terms are independent.

First, we obtain the maximum likelihood estimates for later comparison with the REML estimates.

```
> d$Genotype=factor(d$Genotype)  
>  
> library(lme4)  
> #Maximum Likelihood Estimation  
> lmer(SeedlingWeight~Genotype+(1|Tray), REML=F, data=d)
```



REML = T is  
default setting in R,  
as it is in SAS

Linear mixed model fit by maximum likelihood [‘lmerMod’]

Formula: SeedlingWeight ~ Genotype + (1 | Tray)

Data: d

AIC	BIC	logLik	deviance	df.resid
260.7418	268.8432	-126.3709	252.7418	52

Random effects:

Groups	Name	<u>Std.Dev.</u>	
Tray	(Intercept)	<u>2.932</u>	$\hat{\sigma}_t$
Residual		<u>1.882</u>	$\hat{\sigma}_e$

Number of obs: 56, groups: Tray, 8

Fixed Effects:

(Intercept) Genotype2

15.302 -3.567

Now, we obtain the REML estimates.

Note that REML is the default method for `lmer`.

Although not shown here, REML is also the default for SAS `proc mixed`.

```
> #REML Estimation  
> lmer(SeedlingWeight ~ Genotype + (1 | Tray), data=d)
```

Linear mixed model fit by REML ['lmerMod']

Formula: SeedlingWeight ~ Genotype + (1 | Tray)

Data: d

REML criterion at convergence: 247.1411

Random effects:

Groups	Name	Std.Dev.
Tray	(Intercept)	3.415
	Residual	1.882

Number of obs: 56, groups: Tray, 8

Fixed Effects:

(Intercept)	Genotype2
15.29	-3.55

$\hat{\sigma}_t$  is larger now

compared to its MLE

end  
lecture 35  
4-23-25