20. REML Estimation of Variance Components

Consider the General Linear Model

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

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

- In the previous set of slides, we considered maximum likelihood (ML) estimation of the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

- We saw by example that the MLE of the variance component vector $\boldsymbol{\gamma}$ can be biased.

Example of MLE Bias

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

$$\frac{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n}$$

with expectation

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

This is MLE for σ^2 is often criticized for “failing to account for the loss of degrees of freedom needed to estimate β.”

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

A Familiar Special Case

$$y_1, \ldots, y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$$

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

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

- 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 REsidual Maximum Likelihood or REstricted Maximum Likelihood.

The REML Method

1. Find $n - \text{rank}(\boldsymbol{X}) = n - r$ linearly independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$ such that $\boldsymbol{a}_i'\boldsymbol{X} = \boldsymbol{0}'$ for all $i = 1, \ldots, n - r$.

2. Find the maximum likelihood estimate of γ using $w_1 \equiv \boldsymbol{a}_1'\boldsymbol{y}, \ldots, w_{n-r} \equiv \boldsymbol{a}_{n-r}'\boldsymbol{y}$ as data.

$$\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}] \qquad \boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_{n-r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{y} \\ \vdots \\ \boldsymbol{a}_{n-r}'\boldsymbol{y} \end{bmatrix} = \boldsymbol{A}'\boldsymbol{y}$$

- If $\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{0}', \boldsymbol{a}'\boldsymbol{y}$ is known as an *error contrast*.

- Thus, $w_1, \ldots, w_{n-r}$ comprise a set of $n - r$ error contrasts.

- Because

$$(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \boldsymbol{0},$$

the elements of

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

are each error contrasts.

- Because $\text{rank}(\boldsymbol{I} - \boldsymbol{P_X}) = n - r$, there exists a set of $n - r$ linearly independent rows of $\boldsymbol{I} - \boldsymbol{P_X}$ that can be used in step 1 of the REML method to get $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$.

- If we do use a subset of rows of $\boldsymbol{I} - \boldsymbol{P_X}$ to get $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-r}$; the error contrasts

$$w_1 = \boldsymbol{a}_1'\boldsymbol{y}, \ldots, w_{n-r} = \boldsymbol{a}_{n-r}'\boldsymbol{y}$$

will be a subset of the elements of the residual vector

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

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

Note that

$$\begin{aligned}
\boldsymbol{w} &= \boldsymbol{A}'\boldsymbol{y} \\
&= \boldsymbol{A}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\epsilon} \\
&= \boldsymbol{0} + \boldsymbol{A}'\boldsymbol{\epsilon} \\
&= \boldsymbol{A}'\boldsymbol{\epsilon}
\end{aligned}$$

Thus,

$$\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{\epsilon} \sim N(\boldsymbol{A}'\boldsymbol{0},\ \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}) \overset{d}{=} N(\boldsymbol{0},\ \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}),$$

and the distribution of $\boldsymbol{w}$ depends on γ but not $\boldsymbol{\beta}$.

The log likelihood function in this case is

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

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

In 6II, we take the time to prove that every set of $n - r$ linearly independent error contrasts yields the same REML estimator of γ.

As an example, consider the special case where

$$y_1, \ldots, y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2).$$

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

It follows that

$$\begin{aligned} \boldsymbol{a}_1' &= (1, -1, 0, 0, \ldots, 0) \\ \boldsymbol{a}_2' &= (0, 1, -1, 0, \ldots, 0) \\ &\vdots \\ \boldsymbol{a}_{n-1}' &= (0, 0, \ldots, 0, 1, -1) \end{aligned}$$

and

$$
\begin{aligned}
\boldsymbol{b}_1' &= (1, 0, 0, \ldots, 0, -1) \\
\boldsymbol{b}_2' &= (0, 1, 0, \ldots, 0, -1) \\
&\vdots \\
\boldsymbol{b}_{n-1} &= (0, 0, \ldots, 0, 1, -1)
\end{aligned}
$$

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

Either

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{y} \\ \boldsymbol{a}_2'\boldsymbol{y} \\ \vdots \\ \boldsymbol{a}_{n-1}'\boldsymbol{y} \end{bmatrix} = \begin{bmatrix} y_1 - y_2 \\ y_2 - y_3 \\ \vdots \\ y_{n-1} - y_n \end{bmatrix} \text{ or } \boldsymbol{v} = \begin{bmatrix} \boldsymbol{b}_1'\boldsymbol{y} \\ \boldsymbol{b}_2'\boldsymbol{y} \\ \vdots \\ \boldsymbol{b}_{n-1}'\boldsymbol{y} \end{bmatrix} = \begin{bmatrix} y_1 - y_n \\ y_2 - y_n \\ \vdots \\ y_{n-1} - y_n \end{bmatrix}$$

could be used to obtain the same REML estimator of σ^2.

For the normal theory Gauss-Markov linear model,

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

the REML estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}}{n - r},$$

the unbiased estimator that we used previously.

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.

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

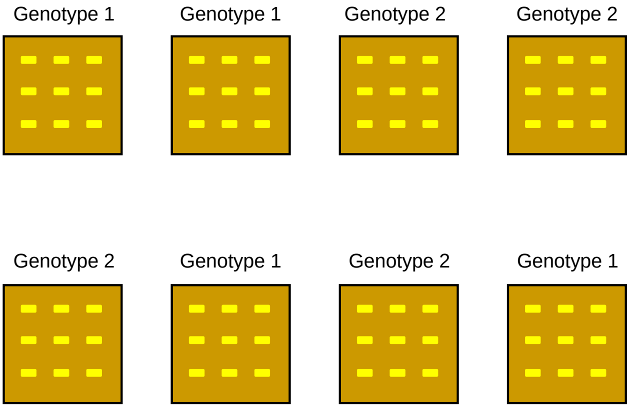
$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}} = \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y},$$

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

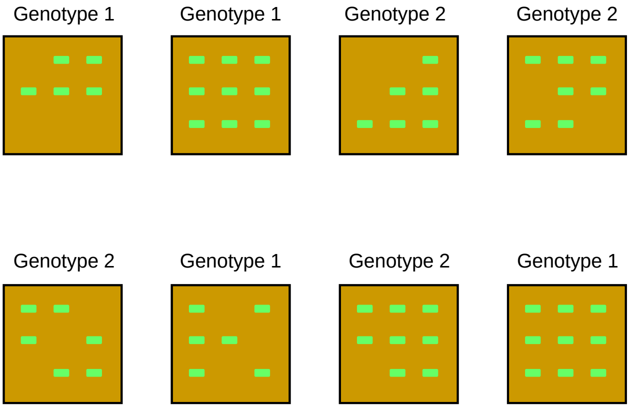
An Example

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

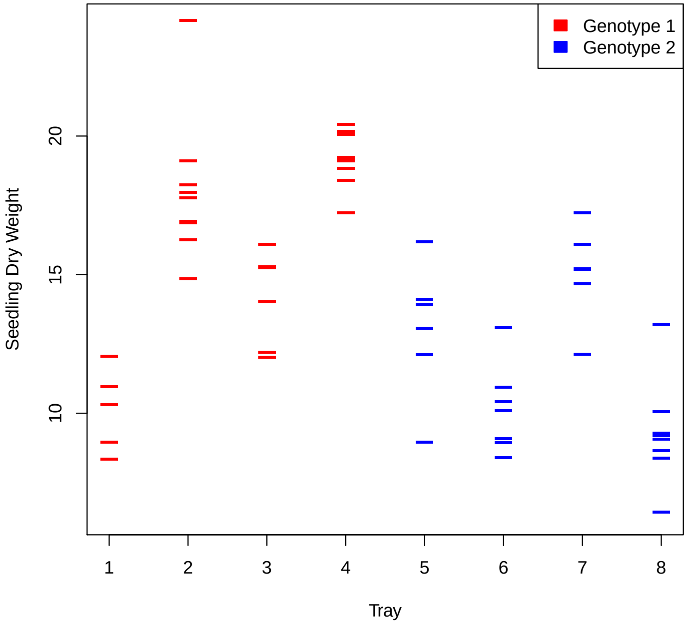


```
> d=read.delim(
+ "https://dnett.github.io/S510/SeedlingDryWeight2.txt")
> d
   Genotype Tray Seedling SeedlingWeight
1         1    1        1              8
2         1    1        2              9
3         1    1        3             11
4         1    1        4             12
5         1    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))
```


Genotype 1
Genotype 2
Seedling Dry Weight
10
15
20
Tray
1
2
3
4
5
6
7
8

A Model for the Seedling Dry Weights

Let y_{ijk} be the dry weight of the kth seedling in the jth tray for genotype i.

Suppose

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

where μ_1 and μ_2 are unknown constants,

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

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(nlme)
>
> lme(SeedlingWeight~Genotype,random=~1|Tray,method="ML",
+     data=d)
```

```
Linear mixed-effects model fit by maximum likelihood
  Data: d
  Log-likelihood: -126.3709
  Fixed: SeedlingWeight ~ Genotype
(Intercept)   Genotype2
  15.301832   -3.567017

Random effects:
 Formula: ~1 | Tray
        (Intercept) Residual
StdDev:    2.932294  1.88247

Number of Observations: 56
Number of Groups: 8
```

```
> library(lme4)
>
> lmer(SeedlingWeight~Genotype+(1|Tray),REML=F,data=d)
```

```
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: SeedlingWeight ~ Genotype + (1 | Tray)
   Data: d
      AIC       BIC    logLik  deviance
 260.7418  268.8432 -126.3709  252.7418
Random effects:
 Groups   Name        Std.Dev.
 Tray     (Intercept) 2.932
 Residual             1.882
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 `lme` and `lmer`.

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

```
> lme(SeedlingWeight~Genotype,random=~1|Tray,data=d)
Linear mixed-effects model fit by REML
  Data: d
  Log-restricted-likelihood: -123.5705
  Fixed: SeedlingWeight ~ Genotype
(Intercept)   Genotype2
  15.288838   -3.550201

Random effects:
 Formula: ~1 | Tray
        (Intercept) Residual
StdDev:    3.414856  1.88223

Number of Observations: 56
Number of Groups: 8
```

```
> 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
```