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 β and γ.

- We saw by example that the MLE of the variance component vector γ can be biased.

Example of MLE Bias

For the case of $\epsilon = \sigma^2\boldsymbol{I}$, where $\gamma = \sigma^2$, 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 $\boldsymbol{\beta}$.”

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

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 Maximun 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 $\boldsymbol{a}_i' X = \boldsymbol{0}'$ 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,

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

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

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

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

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,

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

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.

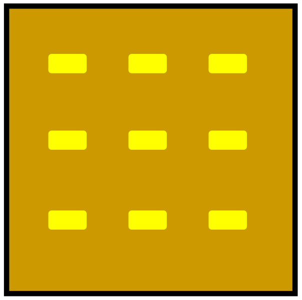
For linear mixed effects models, the REML estimators of variance components produce the same estimates as the unbiased ANOVA-based estimators formed by taking appropriate linear combinations of mean squares when the latter are positive and data are balanced.

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

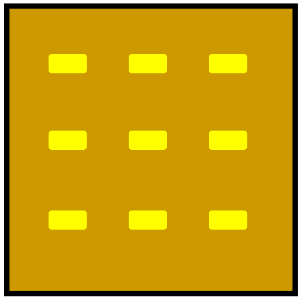
$$\boldsymbol{C}\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}} = \boldsymbol{C}(\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y},$$

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

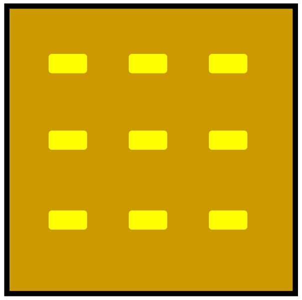
A



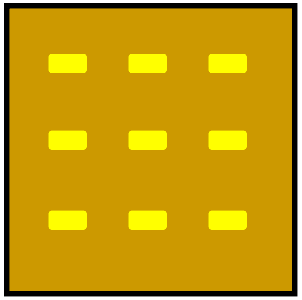
A



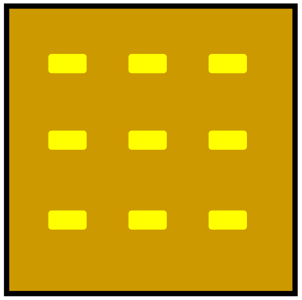
B



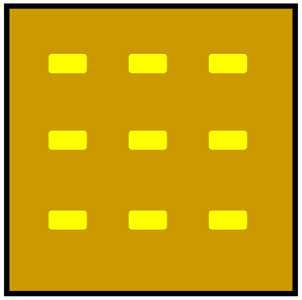
B



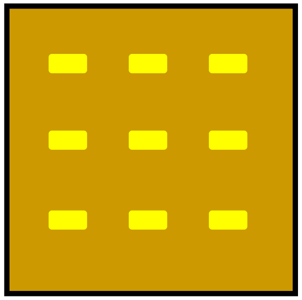
B



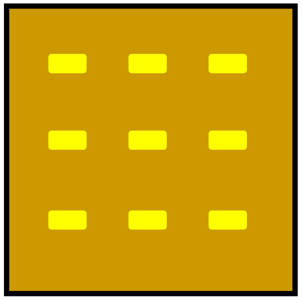
A



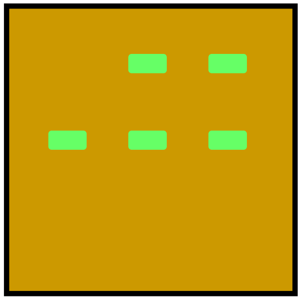
B



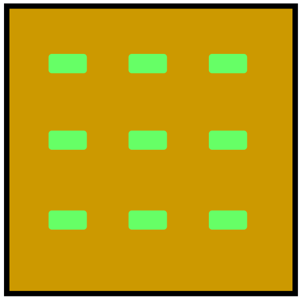
A



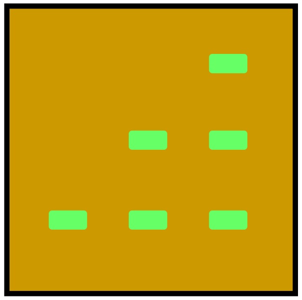
A



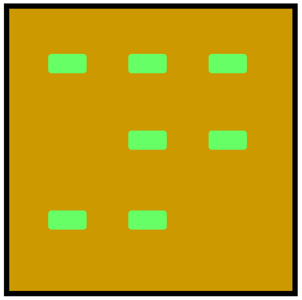
A



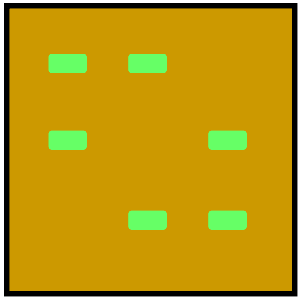
B



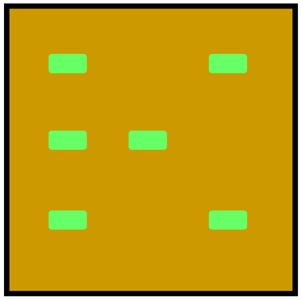
B



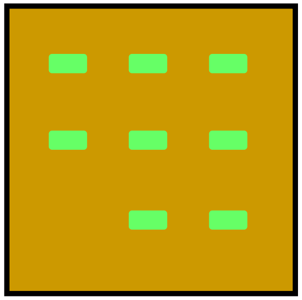
B



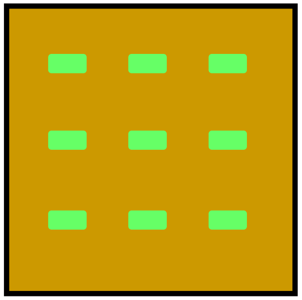
A



B



A



```
d=read.delim(
"http://www.public.iastate.edu/~dnett/S511/SeedlingDryWeight2.txt"
)
d
```

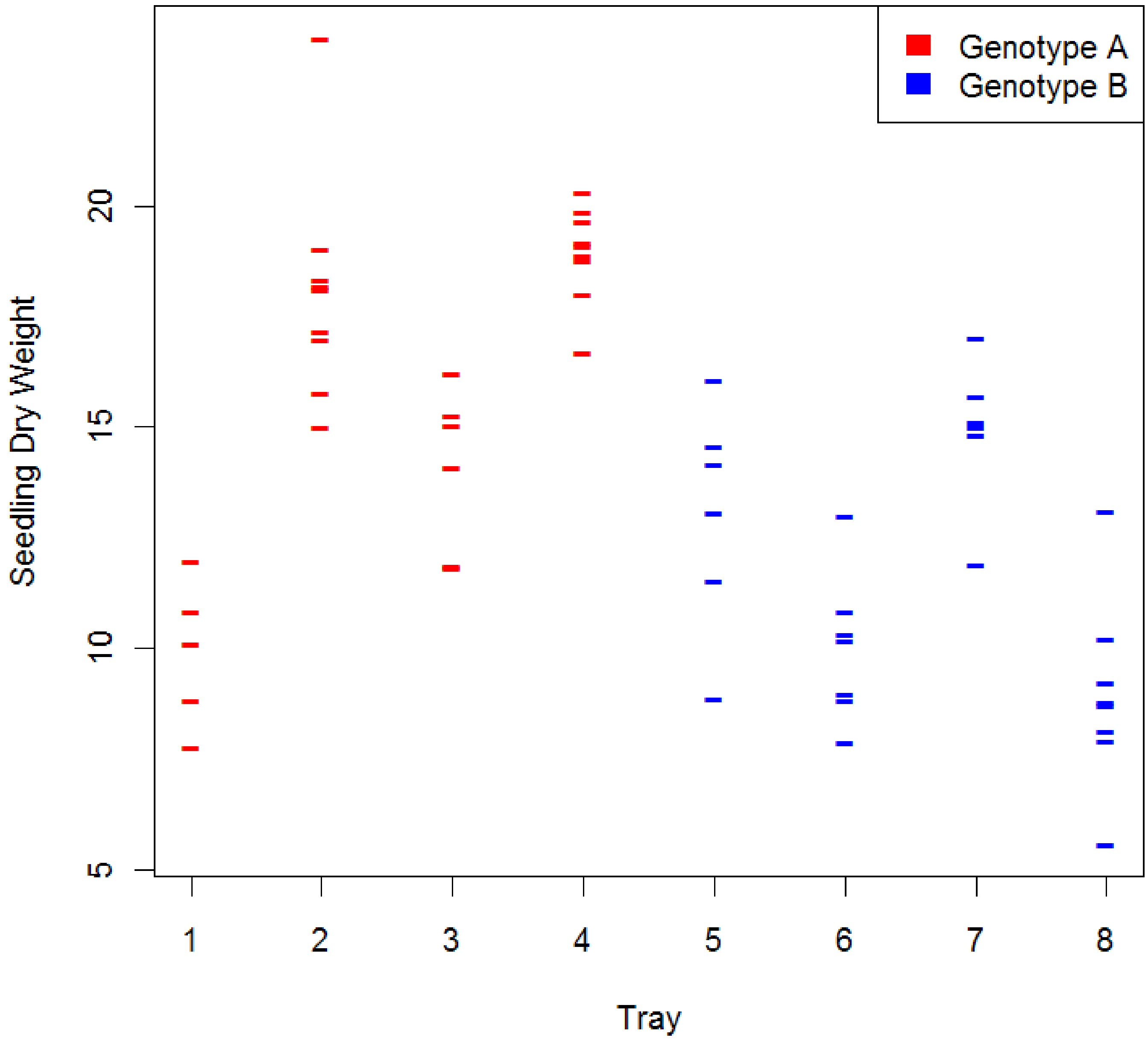
```
   Genotype Tray Seedling SeedlingWeight
1         A    1        1              8
2         A    1        2              9
3         A    1        3             11
4         A    1        4             12
5         A    1        5             10
6         A    2        1             17
7         A    2        2             17
8         A    2        3             16
9         A    2        4             15
10        A    2        5             19
11        A    2        6             18
12        A    2        7             18
13        A    2        8             18
14        A    2        9             24
15        A    3        1             12
16        A    3        2             12
17        A    3        3             16
```

18	A	3	4	15
19	A	3	5	15
20	A	3	6	14
21	A	4	1	17
22	A	4	2	20
23	A	4	3	20
24	A	4	4	19
25	A	4	5	19
26	A	4	6	18
27	A	4	7	20
28	A	4	8	19
29	A	4	9	19
30	B	5	1	9
31	B	5	2	12
32	B	5	3	13
33	B	5	4	16
34	B	5	5	14
35	B	5	6	14
36	B	6	1	10
37	B	6	2	10
38	B	6	3	9

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

```
plot(d[,2],d[,4]+rnorm(56,0,.2),
    xlab="Tray",ylab="Seedling Dry Weight",
    col=2*(1+(d[,1]=="B")),pch="-",cex=2)


legend("topright",c("Genotype A","Genotype B"),
       fill=c(2,4),border=c(2,4))
```



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

```
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)   GenotypeB
  15.301832   -3.567017

Random effects:
 Formula: ~1 | Tray
        (Intercept) Residual
StdDev:    2.932294 1.882470

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

Formula: SeedlingWeight ~ Genotype + (1 | Tray)
   Data: d
   AIC   BIC logLik deviance REMLdev
 260.7 268.8 -126.4    252.7   247.4

Random effects:
 Groups   Name        Variance Std.Dev.
 Tray     (Intercept) 8.5984   2.9323
 Residual             3.5437   1.8825

Number of obs: 56, groups: Tray, 8
```

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)   15.302      1.510   10.14
GenotypeB     -3.567      2.136   -1.67

Correlation of Fixed Effects:
          (Intr)
GenotypeB -0.707
```

```
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)   GenotypeB
  15.288838   -3.550201

Random effects:
 Formula: ~1 | Tray
        (Intercept) Residual
StdDev:    3.414856 1.882230

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

```
lmer(SeedlingWeight~Genotype+(1|Tray),data=d)
```

```
Linear mixed model fit by REML
Formula: SeedlingWeight ~ Genotype + (1 | Tray)
   Data: d
   AIC   BIC logLik deviance REMLdev
 255.1 263.2 -123.6      253   247.1
Random effects:
 Groups   Name        Variance Std.Dev.
 Tray     (Intercept) 11.6612  3.4149
 Residual              3.5428  1.8822
Number of obs: 56, groups: Tray, 8

Fixed effects:
            Estimate Std. Error t value
(Intercept)   15.289      1.745   8.762
GenotypeB     -3.550      2.468  -1.438

Correlation of Fixed Effects:
          (Intr)
GenotypeB -0.707
```