

Lab: Applications for Ordinary Least Squares and Mixed Models

Malachy Campbell

10/16/2018

OLS and MM example 1: Balanced maize data

Learning objectives:

1. Brief overview of ordinary least squares (OLS) and mixed models (MM)
2. Estimate genetic values and h^2 using OLS and MM.
3. Learn to deal with unbalanced data

Maize Dataset

- ▶ 62 recombinant inbred line (RILs) from a cross between B73 and MO17.
- ▶ Randomized complete block design
- ▶ Two replications at four locations
- ▶ Traits: days to pollen, days to silking, anthesis/silking interval (ASI) and plant height.
 - ▶ We'll use height as the response variable.
- ▶ See Isik and

Loading the data.

```
maize <- read.csv("~/Downloads/MaizeRILs.csv")  
  
head(maize)
```

##	location	rep	block	plot	RIL	pollen	silking	ASI	height
## 1	ARC	1	4	28	RIL-1	73	77	4	182.0
## 2	ARC	2	6	47	RIL-1	74	79	5	169.2
## 3	CLY	1	5	36	RIL-1	71	74	3	213.0
## 4	CLY	2	4	223	RIL-1	73	77	4	203.0
## 5	PPAC	1	8	64	RIL-1	97	101	4	155.6
## 6	PPAC	2	5	40	RIL-1	95	100	5	177.6

Obtaining genetic values with OLS

- ▶ For this dataset we can fit the following model:

$$y_{ijk} = \mu + L_i + Rep(L)_{ij} + G_k + GL_{ik} + e_{ijk}$$

- ▶ y_{ijk} is the phenotype (height)
- ▶ L_i is the fixed effect of location i
- ▶ $Rep(L)_{ij}$ is the fixed effect of replicate j nested within location i
- ▶ G_k is the fixed effect of RIL k , GL_{ik} is the interaction of RIL k and location i and e_{ijk} is the residual.

Here's everything except the error term is considered as a fixed effect

Obtaining genetic values with OLS

► Fit the linear model with lm in R

#rep is coded as 1 and 2. So make sure R knows its a factor

```
maize$rep <- as.factor(maize$rep)
```

```
mod1 <- lm(height ~ location*RIL + rep:location, data = maize)
```

```
anova(mod1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: height
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## location	3	84931	28310.4	436.3090	< 2.2e-16 ***
## RIL	61	154938	2540.0	39.1448	< 2.2e-16 ***
## location:RIL	183	20999	114.8	1.7685	1.643e-05 ***
## location:rep	4	3594	898.6	13.8482	3.408e-10 ***
## Residuals	244	15832	64.9		

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtaining genetic values with OLS

- ▶ Use the output of `lm` to estimate the marginal means
- ▶ For RIL-11 we can calculate the marginal means as:

$$RIL11 = \mu + \bar{L} + G_{RIL11} + \bar{G}L_{RIL11} + Rep(\bar{L})$$

Obtaining genetic values with OLS

```
#intercept
MU <- as.numeric(coef(mod1)["(Intercept)"] )
#locations
LOC.eff <- sum(as.numeric(coef(mod1)[c("locationCLY",
    "locationPPAC","locationTPAC")]) )/4
#RIL
RIL1.eff <- as.numeric(coef(mod1)["RILRIL-11"] )
#RIL x Location
RIL1.LOC.eff <- sum(as.numeric(coef(mod1)
    [c("locationCLY:RILRIL-11",
        "locationPPAC:RILRIL-11",
        "locationTPAC:RILRIL-11")]) )/4
#Rep within location
Rep.eff <- sum(as.numeric(coef(mod1)[c("locationARC:rep2",
    "locationCLY:rep2", "locationPPAC:rep2",
    "locationTPAC:rep2")]) )/8

RIL_11 <- MU + LOC.eff + RIL1.eff + RIL1.LOC.eff + Rep.eff

print(RIL_11)
```

```
## [1] 182.875
```


Estimating heritability from ANOVA/OLS

$$h^2 = \frac{\sigma_{RIL}^2}{\sigma_{RIL}^2 + \frac{\sigma_{RIL \times LOC}^2}{n_l} + \frac{\sigma_e^2}{n_r n_l}}$$

► EMS from ANOVA

- Location: $\sigma_e^2 + n_b \sigma_{GL}^2 + n_g \sigma_B^2(L) + n_b n_g \sigma_L^2$
- B(L): $\sigma_e^2 + n_g \sigma_B^2(L)$
- RIL: $\sigma_e^2 + n_b \sigma_{GL}^2 + n_g \sigma_B^2(L)$
- RIL x Loc: $\sigma_e^2 + n_b \sigma_{GL}^2$

Estimating heritability from ANOVA/OLS

- ▶ Since the design is balanced we can estimate H^2 using ANOVA

$$h^2 = \frac{\sigma_{RIL}^2}{\sigma_{RIL}^2 + \frac{\sigma_{RIL \times LOC}^2}{n_l} + \frac{\sigma_e^2}{n_r n_l}}$$

- ▶ $\sigma_{RIL \times LOC}^2 = \frac{MS(RIL \times LOC) - MS(Error)}{n_r}$,
 $\sigma_{RIL}^2 = \frac{MS(RIL) - MS(RIL \times LOC)}{n_r n_l}$, and
 $\sigma_e^2 = MS(Error)$

Estimating heritability from ANOVA/OLS

```
anova.res <- as.data.frame(anova(mod1))

sigma_err <- anova.res[5,3]
sigma_G.E <- (anova.res[3,3] - sigma_err) / 2
sigma_G <- (anova.res[2,3] - anova.res[3,3]) / 8

H2.OLS <- sigma_G / (sigma_G + sigma_G.E/4 + sigma_err/8)
print(H2.OLS)

## [1] 0.9548218
```

Obtaining genetic values (BLUEs) with a mixed model

- ▶ We will fit a mixed model to estimate line values for each RIL
 - ▶ *RIL* as a fixed effect, and *Loc* and *Rep* as random effects
 - ▶ $Var(Loc) \sim N(0, \mathbf{I}\sigma_{LOC}^2)$, $Var(rep) \sim N(0, \mathbf{I}\sigma_{rep}^2)$, and $Var(e) \sim N(0, \mathbf{I}\sigma_e^2)$

Obtaining genetic values (BLUEs) with a mixed model in lme4

- ▶ Random terms are specified by '(1|some term)'.
 - ▶ '(1|location/rep)' is the random effect of rep nested within location
 - ▶ '(1|location:RIL)' is the random effect of location x RIL interaction

```
library(lme4)
mod2 <- lmer(height ~ RIL + (1|location/rep) + (1|location:RIL), maize)

#List the estimates for the fixed effects
summary(mod2)$coefficients[1,1] + summary(mod2)$coefficients[2,1]

## [1] 182.875
```

Estimating heritability with a mixed model in lme4

- Here, all terms with the exception of μ will be considered random.

```
mod3 <- lmer(height ~ 1 + (1|RIL) +  
              (1|location/rep) +  
              (1|location:RIL), maize)  
  
#extract the variance components  
MM.varcomps <- as.data.frame(VarCorr(mod3))  
  
sigma_err.MM <- MM.varcomps[5,4]  
sigma_G.E.MM <- MM.varcomps[1,4]  
sigma_G.MM <- MM.varcomps[2,4]  
  
H2.MM <- sigma_G.MM /  
  (sigma_G.MM + sigma_G.E.MM/4 + sigma_err.MM/8)  
  
print(H2.MM)  
  
## [1] 0.9548218
```

BLUPs for maize height

- ▶ When we want to make a prediction on a random term in the model the predicted value is called BLUP
- ▶ In lme4:

```
mod3 <- lmer(height ~ 1 + (1|RIL) + (1|location/rep)  
             + (1|location:RIL), maize)
```

#extract the blups for RILs

```
blups_m3 <- ranef(mod3)$RIL
```

- ▶ More on BLUPs later!

On your own

- ▶ Run a similar analysis with the unbalanced data and compare OLS and MM approaches
- ▶ Which is more trustworthy?