

Applied Genetic Evaluation - Solution 2

Peter von Rohr

2019-05-06

Problem 1: Analysis of Variance

Estimate the variance component for the sire effect using an analysis of variance. The data is available from https://charlotte-ngs.github.io/GELASMSS2019/ex/w12/data_sire_w12.csv. Because the data contains just female animals, the fixed effect of the sex can no longer be estimated.

Hint

- Use the functions `aov()` to do the analysis of variance and the function `summary()` on the ANOVA result to get the relevant parts of the variance components.

Solution

The data is read using

```
s_data_sire <- "https://charlotte-ngs.github.io/GELASMSS2019/ex/w12/data_sire_w12.csv"
tbl_sire_aov <- readr::read_csv2(file = s_data_sire)
```

```
## Using ',' as decimal and '.' as grouping mark. Use read_delim() for more control.
```

```
## Parsed with column specification:
```

```
## cols(
##   Id = col_double(),
##   slh = col_double(),
##   hrd = col_double(),
##   age = col_double(),
##   cw = col_double(),
##   sire = col_double()
## )
```

```
tbl_sire_aov$slh <- as.factor(tbl_sire_aov$slh)
tbl_sire_aov$hrd <- as.factor(tbl_sire_aov$hrd)
tbl_sire_aov$sire <- as.factor(tbl_sire_aov$sire)
# anova
aov_sire <- aov(cw ~ slh + hrd + age + sire, data = tbl_sire_aov)
(summary_aov_sire <- summary(aov_sire))
```

```
##           Df  Sum Sq Mean Sq F value Pr(>F)
## slh         2  141381   70690   783.9 <2e-16 ***
## hrd         4 1880642  470160  5213.5 <2e-16 ***
## age         1   24152   24152   267.8 <2e-16 ***
## sire        9   10387    1154    12.8 <2e-16 ***
## Residuals 1699  153217     90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the Mean Sq of the residuals and of the sires, the estimates of the variance components can be computed.

The estimate $\hat{\sigma}_e^2$ corresponds to the mean sum of squares for the residuals. Hence

$$\hat{\sigma}_e^2 = 90$$

The estimate of the sire variance $\hat{\sigma}_s^2$ is computed as

$$\hat{\sigma}_s^2 = \frac{\text{Mean Sq(sire)} - \hat{\sigma}_e^2}{k}$$

where $k = \frac{1}{r-1} \left[N - \frac{\sum_{i=1}^r n_i^2}{N} \right]$ with r is the number of sires, N the total number of observations and n_i the number of progeny for sire i . To compute the value k , we need the progeny counts for each sire.

sire	prog_count
1	174
2	189
3	171
4	190
5	176
6	172
7	175
8	160
9	161
10	148

From the above table, we get

$$k = \frac{1}{r-1} \left[N - \frac{\sum_{i=1}^r n_i^2}{N} \right] = 1/(10-1) \left[1716 - \frac{2.95948 \times 10^5}{1716} \right] = 171.5$$

```
n_msqsire <- summary_aov_sire[[1]]$`Mean Sq`[4]
n_hatsigmas2 <- (n_msqsire - n_hatsigmae2) / n_k
```

Putting everything together, we get

$$\hat{\sigma}_s^2 = \frac{1154 - 90}{171.5} = 6.2$$

From the estimate of the sire variance, we get the estimate of the genetic variance by multiplying it with four. Hence

$$\hat{\sigma}_a^2 = 4 * \hat{\sigma}_s^2 = 4 * 6.2 = 24.8$$

Problem 2: Variance Components Estimation Using REML

Use the same data set as for Problem 1 and a sire model to estimate the same sire variance σ_s^2 . The sire model is the linear mixed effects model that contains sire effects as random component. The model can be specified as

$$y = Xb + Zs + e$$

where y is the vector of observations, b is the vector of fixed effects which are the same as in Problem 1, s is the vector of random sire effects and e is the vector of random error terms.

Hint

- Use the package `pedigreemm` to get a REML estimate for the sire variance component σ_s^2 .
- We assume that the sires are not related. Hence variance-covariance matrix $\text{var}(s)$ of the sire components are $\text{var}(s) = I * \sigma_s^2$.

Solution

As the first step, we have to specify a pedigree. The sires are unrelated hence the corresponding pedigree corresponds to

```
ped_sire <- pedigreemm::pedigree(sire = rep(NA, n_nr_sire),
                                dam = rep(NA, n_nr_sire),
                                label = c(1:n_nr_sire))
```

The specified pedigree is used for the linear mixed effects model

```
require(pedigreemm)
```

```
## Loading required package: pedigreemm
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
lmem_sire <- pedigreemm(cw ~ slh + hrd + age + (1|sire),
                       pedigree = list(sire = ped_sire),
                       data = tbl_sire_aov)
summary(lmem_sire)
```

```
## Linear mixed model fit by REML ['lmerpedigreemm']
```

```
## Formula: cw ~ slh + hrd + age + (1 | sire)
```

```
## Data: tbl_sire_aov
```

```
##
```

```
## REML criterion at convergence: 12610.8
```

```
##
```

```
## Scaled residuals:
```

```
##      Min      1Q  Median      3Q      Max
```

```
## -3.2011 -0.6731  0.0137  0.6539  3.4813
```

```
##
```

```
## Random effects:
```

```
## Groups   Name      Variance Std.Dev.
```

```
## sire     (Intercept) 6.257   2.501
```

```
## Residual                90.181  9.496
```

```
## Number of obs: 1716, groups: sire, 10
```

```
##
```

```
## Fixed effects:
```

```
##              Estimate Std. Error t value
```

```
## (Intercept) -77.17270   16.67579  -4.628
```

```
## slh2         22.36751    0.56484  39.600
```

```
## slh3          4.27798    0.56818   7.529
```

```
## hrd2         88.81545    0.73294 121.176
```

```
## hrd3          9.28428    0.72408  12.822
```

```
## hrd4          58.98147    0.71719  82.239
## hrd5          20.36389    0.72889  27.938
## age           0.68269    0.04161  16.405
##
## Correlation of Fixed Effects:
##      (Intr) slh2  slh3  hrd2  hrd3  hrd4  hrd5
## slh2 -0.011
## slh3 -0.072  0.513
## hrd2 -0.018  0.014 -0.023
## hrd3 -0.012 -0.001 -0.003  0.493
## hrd4  0.009 -0.008 -0.025  0.497  0.501
## hrd5  0.004  0.010 -0.028  0.490  0.495  0.500
## age -0.998 -0.007  0.056 -0.004 -0.009 -0.031 -0.026
```

Additional Problem: Variance Components Estimation Using an Animal Model

We are given the dataset with the response variable `carcass weight` (CW) and the predictor variables that resulted from the model selection process from Exercise 1. These consisted of

- sex (`sex`)
- slaughterhouse (`slh`)
- herd (`hrd`)
- age at slaughter (`age`)

The data is available from https://charlotte-ngs.github.io/GELASMSS2019/ex/w12/data_bp_w12.csv.

We use a mixed linear effects model to estimate the variance components for the random effects in the model.

$$y = Xb + Za + e \quad (1)$$

where y is a vector of observations, b is a vector of fixed effects found to be relevant in Exercise 1, a is a vector of random breeding values and e is a vector of random errors.

Hint

- Use the package `pedigreemm` to get an estimate of the variance components

Your Task

- Estimate the variance components σ_a^2 and σ_e^2 for the two random component a and e , respectively.

Solution

We first have to read the data

```
s_data_path_gel_ex2 <- "https://charlotte-ngs.github.io/GELASMSS2019/ex/w12/data_bp_w12.csv"
tbl_gel_ex2 <- readr::read_csv2(file = s_data_path_gel_ex2)
```

```
## Using ',' as decimal and '.' as grouping mark. Use read_delim() for more control.
```

```
## Parsed with column specification:
```

```
## cols(
##   Id = col_double(),
```

```
## sex = col_double(),
## slh = col_double(),
## hrd = col_double(),
## age = col_double(),
## cw = col_double(),
## sire = col_double(),
## dam = col_double()
## )
```

```
colnames(tbl_gel_ex2);dim(tbl_gel_ex2)
```

```
## [1] "Id" "sex" "slh" "hrd" "age" "cw" "sire" "dam"
## [1] 5325 8
```

The fixed effects are converted into factors

```
tbl_gel_ex2$sex <- as.factor(tbl_gel_ex2$sex)
tbl_gel_ex2$slh <- as.factor(tbl_gel_ex2$slh)
tbl_gel_ex2$hrd <- as.factor(tbl_gel_ex2$hrd)
```

From the help file of `pedigreemm`, we can see that we first have to define a pedigree.

```
ped <- pedigreemm::pedigree(sire = tbl_gel_ex2$sire,
                           dam = tbl_gel_ex2$dam,
                           label = tbl_gel_ex2$Id)
```

Now the model can be specified as for the other functions to fit linear mixed effects model, such as `lmer`.

```
# This takes more than one hour to run.
require(pedigreemm)
# according to https://stat.ethz.ch/pipermail/r-sig-mixed-models/2014q1/021609.html
options(lmerControl=list(check.nobs.vs.nlev="ignore",
                        check.nobs.vs.rankZ = "ignore",
                        check.nobs.vs.nRE="ignore"))
s_lmem_file <- "lmem_gel_ex2.rds"
if (file.exists(s_lmem_file)){
  load(file = s_lmem_file)
} else {
  lmem_gel_ex2 <- pedigreemm(cw ~ sex + slh + hrd + age + (1|Id),
                           data = tbl_gel_ex2,
                           pedigree = list(Id = ped))
  saveRDS(lmem_gel_ex2, file = s_lmem_file)
}
summary(lmem_gel_ex2)
```