Project 2 of STA232B:Analyzing the Lamb's Weight Data

Yi Han

Department of Statistics, UC Davis

February 12th, 2020

Abstract

Lamb's weight data[1] shows weights of different lambs and the feature that could influence weights such as lines' type, dams' age and etc. Last project, a linear mixed model was built to depict such relationship. Maximum Likelihood(MLE) and Restricted Maximum Likelihood(RMLE) were used to evaluate parameters. Asymptotic covariance matrix and parametric bootstrap method were used to evaluate the standard error of variance components. In this project, the EBLUP of mixed effects are calculated. Jackknief method is used to estimate the MSPE of EBLUPs.

1 EBLUPs of sire effects

EBLUPs of sire effects are calculated by the following formula:

$$\hat{s} = \hat{\sigma_s}^2 Z'\hat{V}^{-1}(y - X\hat{\beta}) \tag{1}$$

Here, $\hat{\sigma_s}$ is the value of variance of random effects calculated using REML method. $\hat{\beta}$ is the fixed effects calculated by the REML method. Z is the design matrix of random effects. $\hat{V} = ZGZ' + R$ where G, R are variance of random matrix and error variable. Rank the sires according to their EBLUPs, we get the following table(table 1):

Also, we draw a graph of the EBLUPs of all sires(figure 1), the red point represents the largest EBLUP of sire effect, the green point represents the smallest EBLUP of sire effect:

According to the results of EBLUPs of sires, it is easy to draw the conclusion that when considering the lambs' weight, the best sire is 22th sire and the worst is 21th sire. And when we look back as the

Figure 1: Eblups of Sire

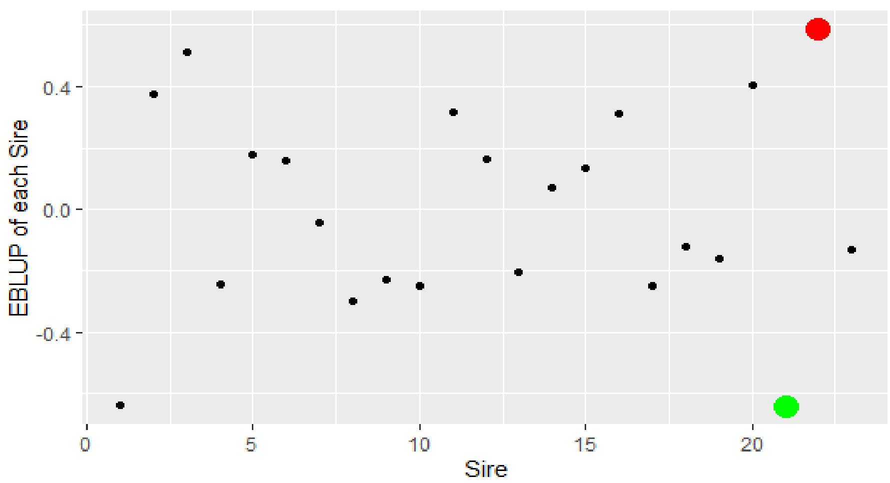


Table 1: EBLUPs of Sire effects

Rank	**Sire**	**EBLUPs**
1	22	0.5856
2	3	0.5112
3	20	0.4041
4	2	0.3732
5	11	0.3144
6	16	0.3104
7	5	0.1776
8	12	0.1626
9	6	0.1613
10	15	0.1336
11	14	0.0714
12	7	-0.0424
13	18	-0.1218
14	23	-0.1299
15	19	-0.1591
16	13	-0.2050
17	9	-0.2287
18	4	-0.2470
19	10	-0.2483
20	17	-0.2486
21	8	-0.2964
22	1	-0.6375
23	21	-0.6408

Table 2: Comparison of EBLUPs

Line	**Sire**	**Age**	**Weight**
5	21	2	10.9
5	21	3	5.9
5	22	2	10.0
5	22	2	12.7
5	22	3	13.2
5	22	3	13.3

lambs' weight data, by putting the EBLUPs of sire together, we get the following table(table 2):

So when sire 21th and 22th are put together with line and age effects, we find that all the Line here are the same: the 5th line. So, we don't need to consider line effects here. When just consider the age effects under the same sire, we can easily observe that when sires are the same, age 2 group behaves better than age 3 group because all the age 2 group has higher birth weight than age 3 group when line and sire are the same. And then we do some in-depth observation of sire 21th and sire 22th. When given the same line 5th and same age group 2, the average weight of lambs whose sire is 21th is 10.9, the average weight of lambs whose sire is 22th is 11.35. When given the same line 5th and same age group 3, the average weight of lambs whose sire is 21th is 5.9, the average weight of lambs whose sire is 22th is 13.25. So, it is obvious that sire 22th behaves much better than 21th in the sense of producing lambs with more weight.

Table 3: EBLUPs of line-wise mean

Rank	Line	EBLUP
1	2	12.300
2	5	10.953
3	3	10.900
4	1	10.690
5	4	10.200

2 EBLUps of line-wise mean

Here, the line wise mean is defined as the following[2]:

$$L_i = E(\frac{1}{N_i}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} y_{ijk}|s) \tag{2}$$

Where $N_i = \sum_{j=1}^{n_i} n_{ij}$ and $s = (s_{ij})_{1i\leq 5, 1j\leq n_i}$, we can rewrite it as:

$$L_i = E(\frac{1}{N_i}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} y_{ijk}|s) = M_i x\beta + M_i Zs =: a'\beta + b'\beta s \tag{3}$$

So, we can calculate the EBLUPs of L_i by substituting the β and s with $\hat{\beta}$ and $\hat{s}$, which is the REML estimator of β and s.

The result shows in the following table(table3):

The line-wise mean can be actually considered as a mixed effect, then the estimated line-wise mean are just EBLUP of mixed effects. Combining the fixed effect and age effects in line i, the EBLUP of L_i is the EBP of average weight of lambs in line i. According to the result in table 6, we can easily find that lambs from line 2 have the heaviest weight.

Comparing with the previous analysis of REML of fixed effects, the ranking of line effects are very similar to the ranking of line-wise mean here. It is very interesting because it may suggest that line effects play the major role in fixed effects when compared with age effects. And this is very reasonable when considering the model selection we do last project, it also shows that the age effect are not significant under the 0.05 significance level.

3 JLW jackknife estimate of MSPE

Using the JLW jackknife estimation method to estimate the MSPE of EBLUP of the sire effect for part 1. The formula show below:

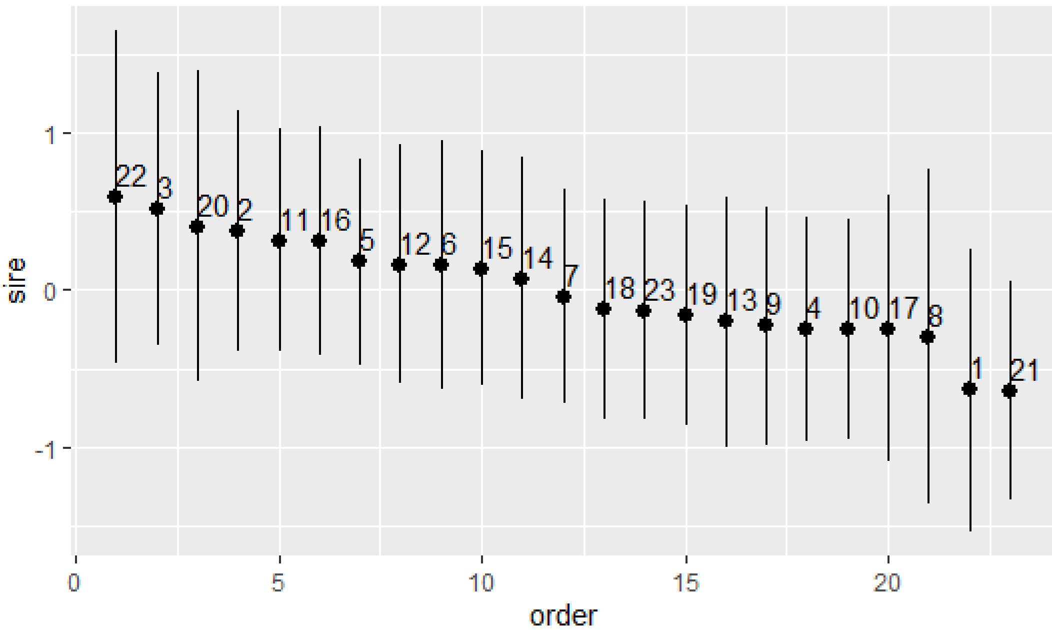
$$MSPE(\hat{s_{ij}}) = b_{ij}(\hat{\phi}) - \frac{m-1}{m}\sum_{i',j'}\{b_{ij}(\hat{\phi}_{-(i',j')}) - b_{ij}(\hat{\phi})\} + \frac{m-1}{m}\sum_{(i',j')}\{\hat{s}_{ij,-(i',j')} - \hat{s_{ij}}\}^2 \tag{4}$$

where $\hat{\phi} = (\hat{\beta}, \hat{\sigma_s}^2, \hat{\sigma_e}^2)$, $b_{ij}(\hat{\phi}) = \frac{\sigma_e^2\sigma_s^2}{\sigma_e^2 + n_{ij}\sigma_s^2}$, $\hat{s} = \hat{\sigma}_s^2 Z'\hat{V}^{-1}(y - X\hat{\beta})$. $\hat{s}_{ij} = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_e^2 + n_{ij}\hat{\sigma}_s^2}\sum_{k=1}^{n_{ij}}(y_{ijk} - x'_{ijk}\hat{\beta})$.

So, use the formula above, we can calculate the MSPE of each $\hat{s}_{ij}$, which is the MSPE of EBLUP of each sire effect. Here, I construct a 23×23 matrix whose j-th column represents the data deleting the $(i,j)th$ data; i-th row represent the MSPE of $(i,j)th$ EBLUP. And square root of the MSPE are set as the "margin of error". The graphic of mspe of each sire's EBLUP are shown in the following graphic:

The graph shows the EBLUP of 23 sire and their margin od error. The black point is their EBLUPs and the line below and above the black points represent the margin of error of each sire. From the graph

Figure 2: MSPE of Sire's EBLUP



we can draw the conclusion that when the EBLUPs are relatively larger or smaller than 0, the margin of error will also be larger than other sires. The reason may be that it is quite difficult to make prediction about extreme values, increasing the error of such prediction.

4 Discussion

From the analysis, we can draw the conclusion that the 22th sire has the highest EBLUP and relatively high MSPE. And the 3rd sire has a little smaller EBLUP but also relatively lower MSPE, the margin of error. So in practice, I recommend farmers use 3rd sire as a good ram for helping produce lambs with heavier born weight to make more profits.

References

[1] Harville D A, Fenech A P. Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model[J]. *Biometrics*, 1985, 41(1): 137-152.

[2] Jiang, Jiming. *Linear and generalized linear mixed models and their applications*. Springer Science Business Media, 2007.