HW2

Neal Marquez

January 23rd, 2019

Homework 2

This homework uses data from the 1975 U.S. Sustaining Effects Study of elementary education, available as the egsingle dataset in the `mlmRev` R package. This gives data on 1,721 students in 60 schools. We will take `math` (Mathematics achievement score) as the outcome variable, and `schoolid` (the code for the school the student attends) as the grouping variable. For this homework, we will use the data for year 0.5 only.

Question 1

In this question we will consider the one-way random effects analysis of variance model. We will do many of the same things we did in Homework 1 for the classical analysis of variance model, and compare the results.

**A. Write out the model formally.**

$$y_i = \alpha_{j[i]} + \epsilon_i, \ i = 1, \dots, n$$
$$\epsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_y^2)$$
$$\alpha_j \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2) \ j = 1, \dots, J$$

B. Estimate the parameters of the model.

We may estimate the parameters of the model by using a variant of maximum likelihood which is formulated in the `lme4` package in R.

C. Plot the model parameters and their confidence intervals in informative ways.

School Effects As Estimated by Random Effects Model(w/ 95% CI)
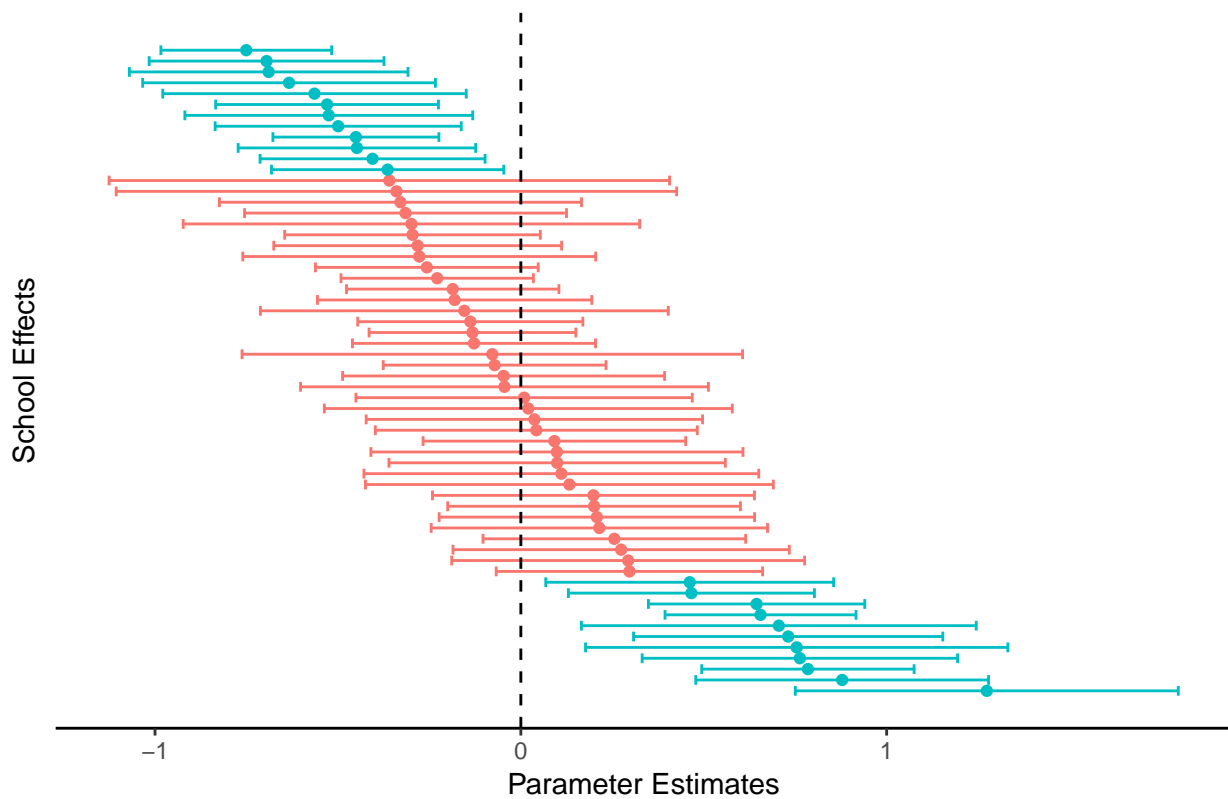
D. Carry out a formal statistical test for the significance of the school effect.

As stated in Bates et al. 2015 we may use an anova test to measure the inclusion of random effects against a base model that matches the original models fixed effects and only does not include the random effects as long as we fit our random effects model with `REML=FALSE`. The Null in this case tells us that the school level effects are null or more formally $\sigma_\alpha^2 = 0$ whereas the alternative hypothesis states that there is a significant difference between the groups or $\sigma_\alpha^2 \neq 0$. To We can test this in `R` by building a null model that is a linear model with only an intercept, $\hat{\mu}_\alpha$, and residual variance, $\hat{\sigma}_y^2$, and comparing it to the the random effects model described above. This is done in the anova with the χ^2-test which will have a single degree of freedom because of the additional variance parameter estimated σ_α^2. The output below shows that the inclusion of random effects provides us with a significantly better fit or that we should accept the alternative hypothesis at the $\alpha = .05$ level, with a p-value of $< 2.2e{-}16$.

```
## Data: schoolDF
## Models:
## LM0: math ~ 1
## M0: math ~ 1 + (1 | schoolid)
##     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## LM0  2 5367.8 5378.6 -2681.9   5363.8
## M0   3 5166.8 5183.1 -2580.4   5160.8 202.97      1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

E. Estimate the intraclass correlation coefficient.

We may use the estimates of σ_α^2 and σ_y^2 from the random effects model directly to get an estimate of the interclass correlation coefficient. We estimate $\hat{\sigma}_\alpha^2$ to be 0.2466 while $\hat{\sigma}_y^2$ is 1.204 giving us an estimate of the

interclass correlation coefficient, $\hat{\rho}$, to be 0.17.

**F. What conclusions do you draw from this analysis?**

We may conclude that accounting for the effects of schools on the outcome of an individual students math score has a significant effect and tells us more than we would know if we did not have the school information. Though the inter class correlation coefficient is relatively small an anova testing the inclusion of random effects showed that the inclusion of school level random effects provided a significantly better fit to the outcome.

Question 2

We will now compare the results from the classical analysis of variance model with the one-way random effects analysis of variance model.

A. Plot the estimated school means from the random effects model against those from the classical model and comment on the results.

The plot below shows a comparison for the school effects as estimated by random effects and fixed effects models. We can immediately see that the estimates from the random effects model tend to be pulled toward the mean especially for schools where the uncertainty is large, i.e. there are not many student observations. Another consequence of this is that the individual school effects uncertainty are smaller as well.
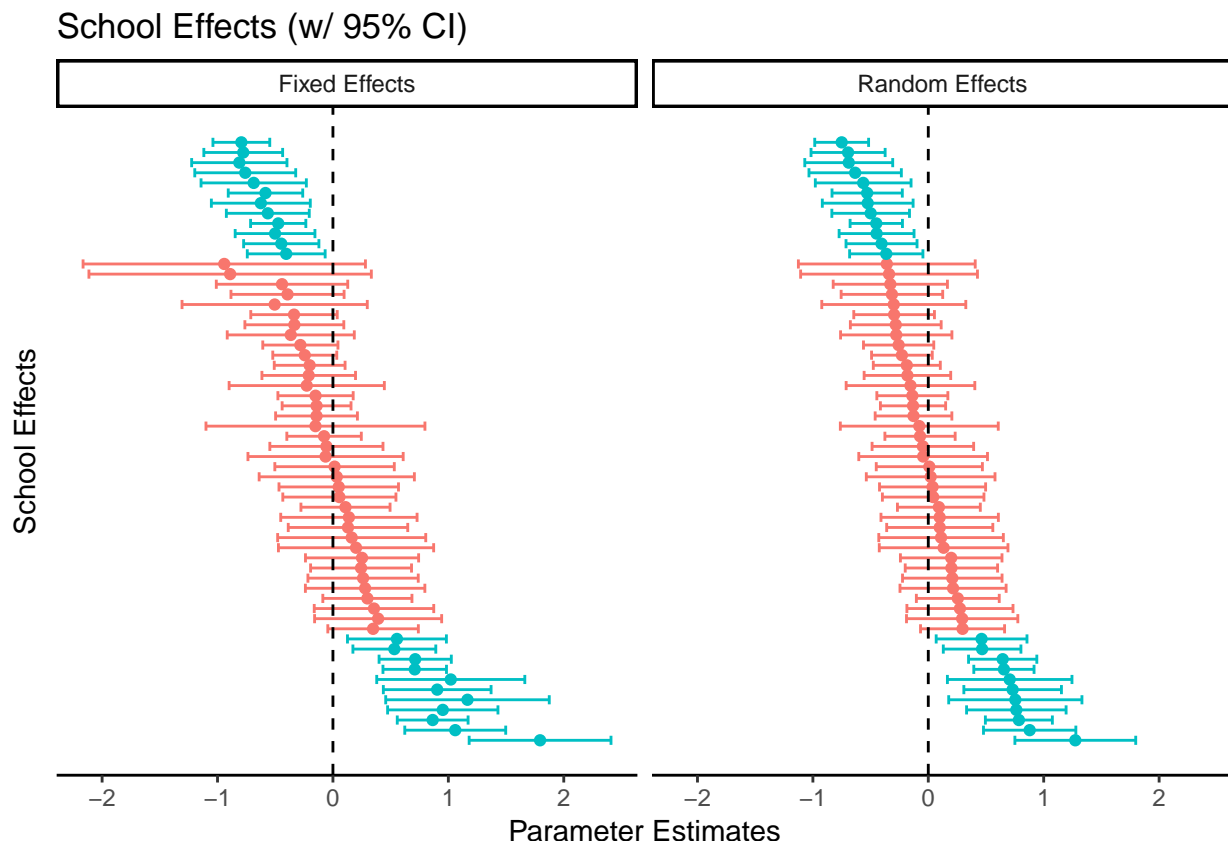
B. For which school is the difference between the two estimates the largest?

The biggest change occurred for `schoolid` 3221 with a change in score of 0.5823. This is to be expected as the number of student observations that occurred for this school was 3 which means that its observational

weight was much smaller compared to the weight of the overall mean, leading to an estimate in the random effects model that is closer to zero.

C. For this one school, calculate (i) the overall mean; (ii) the school mean; (iii) the weights for the overall mean and the school mean in forming the multilevel estimate of the school mean. (iv) Use these results to write the multilevel estimate of the school mean explicitly as a weighted average, giving the numerical values, and compare your result with the result from lmer software.

In the following notation I refer to the school that had the greatest change in the school level estimate as $j^\star$. We may calculate the school mean of $j^\star$ by simply adding all math score observations that take place within that school and dividing by the number of observations at that school.

$$\bar{y}_{j^\star} = \frac{\sum_{i \text{ where } j[i]=j^\star} y_i}{\sum_{i \text{ where } j[i]=j^\star} 1}$$

Our school mean is then found to be -1.2557.

We can use each school mean to calculate an estimate of the overall mean $\hat{\mu}_\alpha$ in the following way.

$$\hat{\mu}_\alpha = \frac{\sum_{j=1}^{J} c_j \bar{y}_j}{\sum_{j=1}^{J} c_j}$$

$$c_j = \frac{1}{\sigma_\alpha^2 + \frac{\sigma_y^2}{n_j}}$$

The estimate for the overall mean is then -0.3113.

To estimate the random effect of $j^\star$ we must calculate weights $\text{w}_{1j^\star}$ and $\text{w}_{2j^\star}$ which are found to be 2.4918 and 4.0547 respectively calculated from the equations.

$$\text{w}_{1j^\star} = \frac{n_{j^\star}}{\sigma_y^2} \text{w}_{2j^\star} = \frac{1}{\sigma_\alpha^2}$$

Finally, we may get the school level random effect $\hat{\alpha}_{j^\star}$ from the equation

$$\hat{\alpha}_{j^\star} = \frac{\text{w}_{1j^\star} \bar{y}_{j^\star} + \text{w}_{2j^\star} \hat{\mu}_\alpha}{\text{w}_{1j^\star} + \text{w}_{2j^\star}}$$

The final estimate for the school level random effect from this hand calculation is found to be -0.6707 which is exactly the same as the sum of the intercept and $j^\star$ random effect from the `lme4` output.

D. What are the main factors contributing to the difference between the classical analysis of variance estimates and the random effects analysis of variance estimates?

The main factors that contributes to the difference is the sample size of students from the $j^\star$ school, leading to a smaller weight calculation for the group mean relative to the overall mean, and how random effects take into account both the group mean and overall mean in the school effect where is fixed effects are only concerned with the group mean for the mean of the fixed effect estimate.

References

Bates, et. al. 2015. "Fitting Linear Mixed-Effects Models Using lme4". Journal of Statistical Software. October Vol 67, Issue 1.