Variance component estimation for the one-way random effects model

Stat 612 Final report (Spring 2019)

Yewon Kim, Reetam Majumder, Mark Ramos

Department of Mathematics and Statistics, UMBC

Abstract

We consider various methods to generate confidence intervals for the variance component from a one-way random effects model in balanced and unbalanced cases. The primary focus is on one-sided lower confidence bounds. Multiple bootstrap methods, Generalized Confidence Intervals and some approximate methods based on the ratio of the variance components. Non-parametric bootstrap had sharper bounds but low coverage, while parametric bootstraps were very conservative. The other methods provided a better balance of results. Possible improvements to existing approaches and other methods that can be compared are laid out as next steps.

1 Introduction

One of the most common inference problems involves the one-way random effects model and the estimation of its variance components. While the approach for the balanced data case is relatively straightforward, there does not seem to be consensus regarding the best way to estimate the between-group variance in the unbalanced case. Further, the definition of *unbalanced* itself can vary. In this report, we have considered balance in terms of the variance of the group sizes. The variance is minimized if each group has the same number of observations, and maximized if exactly 2 of the groups have a much higher group size than the remaining groups. Every other configuration would have variance between these two extremes. Of course, other definitions of balance may be constructed, but this is one of the easiest to visualize and employ.

Bootstrapping is an obvious candidate for tackling this problem. There are however multiple ways to generate bootstrap estimates from the given data. Similarly, Generalized Confidence Intervals is another approach to estimate the between-group variance. There are also other approximate intervals that can be generated by using the Thomas-Hultquist and Burdick-Eickman methods. Many of these can be improved using bias correction methods.

Section 2 introduces the model and the simulation algorithm used to generated data. Section 3 discusses the methodologies used to compute the confidence intervals. Section 4 provides results pertaining to coverage and average interval length for the different methods. Section 5 discusses possible improvements and future work using other estimation methods.

2 Model and data

The data was generated using `R` from the following model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad j = 1, 2, \ldots, n_i,$$
$$i = 1, 2, \ldots, k$$

where $\mu = 10$ is the common mean, $\tau_i \sim N(0, \sigma_\tau^2)$ is the treatment effect and $\epsilon_{ij} \sim N(0, \sigma^2)$ is the noise. We considered $k = 10$ & 20, $\sigma_\tau^2 = 0.1, 1,$ & 2. $N = \sum_i n_i$ is set as $10 \times k$. Three levels of imbalance are considered. The balanced case was set as $n_i = 10$ for all i. In the slightly unbalanced case, n_1 and n_2 were set to be 5 more than the other n_i's. In the highly unbalanced case, each group was set to have only 5 samples except for two groups which shared the remainder of the sames (for k=10, $n_1 = n_2 = 30$; for k=20, $n_1 = n_2 = 55$).

3 Methodology

3.1 Bootstrap lower confidence bounds

We have considered 3 different bootstrapping methods - parametric bootstrap, nonparametric bootstrap, and a bootstrap based on model residuals. For all the three algorithms, the output is a single lower confidence bound for σ_τ^2. To get the coverage for the confidence interval, we repeat this procedure for multiple iterations, getting multiple confidence bounds. The percentage of confidence bounds containing the true value of σ_τ^2 is the coverage of the coverage of the confidence interval, and the average value of the lower bound in these cases is the average length of the confidence bound. In our simulations, $M = 1000$ and the coverage is calculated based on 200 iterations.

Algorithm 1 Generate a lower CI for σ_τ^2 using non-parametric bootstrapping

```
for m in 1 to M do                                    ▷ M bootstrap samples from the original sample
    for i in 1 to k do                                ▷ For each of the k groups
        Sample y*_{i1}, ... y*_{in_i} with replacement from y_{i1}, ... y_{in_i}
    Fit a one-way random effects model to y*_{11}, ... y*_{kn_k}
    Estimate σ*²_{τ,m} from the model
Get the empirical distribution function of σ*²_τ based on σ*²_{τ,1} ... σ*²_{τ,M}
Use the upper α percentile to get the lower CI for σ*²_τ as (σ*²_{τ,1−α}, ∞)
```

Algorithm 2 Generate a lower CI for σ_τ^2 using parametric bootstrapping

Get estimates $\hat{\sigma}^2, \hat{\sigma_\tau}^2$ & $\hat{\mu}$ based on $y_{11}, \ldots, y_{kn_k}$
for m in 1 to M **do** ▷ M bootstrap samples from the original sample
  Sample σ_m^{*2} from $\frac{\hat{\sigma}^2 \times \chi^2_{N-k}}{N-k}$
  Generate $y^*_{11}, \ldots, y^*_{kn_k}$ using $\hat{\mu}, \hat{\sigma_\tau}^2$, and σ_m^{*2}
  Fit a one-way random effects model to $y^*_{11}, \ldots y^*_{kn_k}$
  Estimate $\sigma^{*2}_{\tau,m}$ from the model
Get the empirical distribution function of σ_τ^{*2} based on $\sigma^{*2}_{\tau,1} \ldots \sigma^{*2}_{\tau,M}$
Use the upper α percentile to get the lower CI for σ_τ^{*2} as $(\sigma^{*2}_{\tau,1-\alpha}, \infty)$

Algorithm 3 Generate a lower CI for σ_τ^2 using residual bootstrapping

if method BLUE **then**
  Use one-way fixed effects model to estimate $\hat{\mu}, \hat{\tau}_1, \ldots, \hat{\tau}_k$, and residuals $e_{11}, \ldots, e_{kn_k}$
else if method BLUP **then**
  Use one-way random effects model to estimate $\hat{\mu}, \hat{\tau}_1, \ldots, \hat{\tau}_k$, and residuals $e_{11}, \ldots, e_{kn_k}$
for m in 1 to M **do** ▷ M bootstrap samples from the original sample
  for i in 1 to k **do** ▷ For each of the k groups
    Sample $e^*_{i1}, \ldots e^*_{in_i}$ with replacement from $e_{11}, \ldots e_{kn_k}$
    Sample τ^*_i from $\hat{\tau}_1, \ldots, \hat{\tau}_k$
    Compute $y^*_{ij} = \hat{\mu} + \tau^*_i + e^*_{ij}, \quad j = 1, \ldots, n_i$
  Fit a one-way random effects model to $y^*_{11}, \ldots y^*_{kn_k}$
  Estimate $\sigma^{*2}_{\tau,m}$ from the model
Get the empirical distribution function of σ_τ^{*2} based on $\sigma^{*2}_{\tau,1} \ldots \sigma^{*2}_{\tau,M}$
Use the upper α percentile to get the lower CI for σ_τ^{*2} as $(\sigma^{*2}_{\tau,1-\alpha}, \infty)$

3.2 Generalized Confidence Intervals

GCI is a method for constructing confidence intervals for a statistic using the generalized pivot $g(T(x, X, \theta_i))$, where x is the actual realization, X is the random variable and θ_i's are the parameters of interest. T has a known distribution when conditioned on the data and $T(X = x) = \theta_i$ for each i. In the context of this study, the estimator of interest is $\hat{\sigma_\tau}^2 = \frac{MST - MSE}{n_0}$ which is known to be approximately a linear combination of two χ^2 random variables. Based on this, it was derived that

$$g(T_i) = \frac{(MST(k-1)/n_0)}{\chi^2_{k-1}} - \frac{((N-k)MSE/n_0)}{\chi^2_{N-k}}$$
$$n_0 = \frac{1}{k-1}(N - \frac{\sum_i n_i^2}{N})$$

The algorithm for computing generalized confidence intervals that was used is described in the following steps.

Algorithm 4 Generate a Generalized Confidence Interval for σ_τ^2

1: Compute MSE, MST and n_0 using method of moments from $y_{11}, \ldots, y_{kn_k}$
2: **for** m in 1 to M **do**
3: Generate from χ^2_{k-1} and χ^2_{N-k}
4: Compute a single sample of $\hat{\sigma_\tau}^2$ from $g(T_i)$
5: Obtain the confidence interval for σ_τ^2 using the corresponding percentiles from the generated samples

In this example, $M = 10000$ and 1000 iterations used to calculate the coverage and average interval length.

3.3 Approximate confidence intervals

Consider the quantities

$$MS2 = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y_{i.}})^2,$$

$$MS3 = \frac{1}{k-1}\sum_{i=1}^{k}(\bar{y_{i.}} - \frac{1}{k}\sum_{i=1}^{k}\bar{y_{i.}})^2$$

Thomas and Hultquist (1978) showed that it holds approximately:

$$\frac{(k-1)MS3}{\sigma_\tau^2 + \sigma^2/\tilde{n}} \sim \chi^2_{k-1}$$

where $\tilde{n}$ is the harmonic mean of the sample sizes of the k groups. Further, they proved that MS2 and MS3 are stochastically independent and

$$\frac{\sigma^2}{\sigma_\tau^2 + \sigma^2/\tilde{n}} \cdot \frac{MS3}{MS2} \sim F_{k-1,N-k}$$

approximately. Based on this, they proposed the following confidence interval for σ_τ^2:

$$CI_{TH}(\sigma_\tau^2) : [\frac{k-1}{\chi^2_{k-1;1-\alpha/2}}(MS3 - \frac{MS2}{\tilde{n}}F_{k-1,N-k;1-\alpha/2}); \frac{k-1}{\chi^2_{k-1;\alpha/2}}(MS3 - \frac{MS2}{\tilde{n}}F_{k-1,N-k;\alpha/2})]$$

However, they also reported that the χ^2-approximation is not good for extremely unbalanced designs where the ratio $\eta = \sigma_\tau^2/\sigma^2$ is less than 0.25. In such cases the CI can be quite liberal. Burdick and Eickman (1986) suggested an alternate confidence interval:

$$CI_{BE}(\sigma_\tau^2) : [(\frac{\tilde{N}L}{1+\tilde{n}L}) \cdot \frac{(k-1)MS3}{\chi^2_{k-1;1-\alpha/2}}; (\frac{\tilde{N}U}{1+\tilde{n}U}) \cdot \frac{(k-1)MS3}{\chi^2_{k-1;\alpha/2}}]$$

with

$$L = max[0, \frac{MS3}{MS2} \cdot \frac{1}{F_{k-1,N-k;1-\alpha/2}} - \frac{1}{n_{min}}]$$
$$U = max[0, \frac{MS3}{MS2} \cdot \frac{1}{F_{k-1,N-k;\alpha/2}} - \frac{1}{n_{max}}]$$

Both these methods have been modified to provide one-sided confidence bounds of level $\alpha = 0.05$. Coverage was calculated as in the previous cases, by averaging over 1000 iterations.

4 Results

Table 4.1: Generalized confidence intervals with coverage for σ_τ^2

Setting	σ_τ^2	k	b	1 sided coverage	2 sided coverage	1 sided length	2 sided length
1	0.1	10	1	0.953	0.938	0.0383	0.4261
2	0.1	10	2	0.967	0.953	0.0324	0.3794
3	0.1	10	3	0.967	0.938	0.0330	0.3900
4	1	10	1	0.955	0.945	0.4776	2.9412
5	1	10	2	0.978	0.955	0.4101	2.6506
6	1	10	3	0.964	0.915	0.3614	2.5475
7	2	10	1	0.944	0.946	0.9761	5.7705
8	2	10	2	0.983	0.96	0.8518	5.1923
9	2	10	3	0.958	0.907	0.7669	4.9899
10	0.1	20	1	0.959	0.962	0.0422	0.2311
11	0.1	20	2	0.961	0.959	0.0407	0.2225
12	0.1	20	3	0.935	0.902	0.0366	0.2286
13	1	20	1	0.948	0.943	0.5850	1.6147
14	1	20	2	0.967	0.943	0.5438	1.5413
15	1	20	3	0.911	0.845	0.4901	1.4978
16	2	20	1	0.953	0.947	1.1881	3.1759
17	2	20	2	0.973	0.953	1.1083	2.9911
18	2	20	3	0.914	0.849	0.9832	2.8413

Table 4.1 summarizes the results conducted on simulations using the generalized confidence interval method. Coverage for the balanced and slightly unbalanced cases (b=1,2) was acceptable (around 95 percent) across all settings. Only in the very unbalanced case (b=3) did coverage diminish considerably past acceptable levels for the two-sided case when number of groups k=10 and for both one-sided and two-sided cases when k=20. Correspondingly, there were slight improvements observed on the interval lengths in both one-sided and two-sided cases when the data was very unbalanced.

Table 4.2: Average lower confidence bound for σ_τ^2 with coverage based on non-parametric and parametric bootstrapping

				Non-parametric		Parametric	
Setting	σ_τ^2	k	b	Coverage	Length	Coverage	Length
1	0.1	10	1	0.745	0.0445	1	0.0157
2	0.1	10	2	0.84	0.0425	1	0.0056
3	0.1	10	3	0.82	0.0331	1	0.0038
4	1	10	1	0.695	0.6331	1	0.3533
5	1	10	2	0.795	0.5865	1	0.2429
6	1	10	3	0.795	0.5419	1	0.2199
7	2	10	1	0.645	1.2952	1	0.7390
8	2	10	2	0.755	1.2105	1	0.5244
9	2	10	3	0.765	1.1377	1	0.4979
10	0.1	20	1	0.645	0.0683	0.99	0.0312
11	0.1	20	2	0.695	0.0644	0.995	0.0250
12	0.1	20	3	0.67	0.0597	0.995	0.0140
13	1	20	1	0.65	0.7022	0.99	0.5033
14	1	20	2	0.76	0.6786	1	0.4228
15	1	20	3	0.705	0.7020	0.99	0.4410
16	2	20	1	0.63	1.4194	0.99	1.0273
17	2	20	2	0.74	1.3807	1	0.8685
18	2	20	3	0.705	1.4570	0.99	0.9253

Table 4.3: Average lower confidence bound for σ_τ^2 with coverage based on BLUP and BLUE based residual bootstrapping

				BLUP estimation		BLUE estimation	
Setting	σ_τ^2	k	b	Coverage	Length	Coverage	Length
1	0.1	10	1	1	0.0087	1	0.0063
2	0.1	10	2	1	0.0046	1	0.0050
3	0.1	10	3	1	0.0023	0.995	0.0017
4	1	10	1	0.99	0.3024	0.995	0.3036
5	1	10	2	0.995	0.2431	0.995	0.2386
6	1	10	3	0.995	0.2014	1	0.2134
7	2	10	1	0.99	0.6534	1	0.6606
8	2	10	2	0.995	0.5338	0.995	0.5202
9	2	10	3	0.995	0.4754	0.995	0.5177
10	0.1	20	1	0.995	0.0161	1	0.0168
11	0.1	20	2	0.995	0.0142	1	0.0127
12	0.1	20	3	1	0.0054	1	0.0066
13	1	20	1	0.995	0.4744	0.99	0.4713
14	1	20	2	1	0.4109	0.995	0.4088
15	1	20	3	0.995	0.4078	0.99	0.3988
16	2	20	1	0.995	0.9927	0.99	0.9891
17	2	20	2	1	0.8657	0.995	0.8601
18	2	20	3	0.995	0.8941	0.99	0.8814

Tables 4.2 and 4.3 show the results from the various bootstrap algorithms. Firstly, note than in Table 4.3 there does not seem to be a consistent better performer between the BLUP and BLUE approaches of model residual bootstrapping. We will henceforth refer to the BLUP method when talking about residual bootstrap. Next, note that both the parametric approach and the residual approach (which is also a form of parametric approach) sacrifice interval size while having over 99% in all cases. Non-parametric bootstrap on the other hand has much less coverage but a shorter interval than both other bootstrap methods and even GCI (Table 4.1) in all one-sided cases. Thus it tends to overestimate the variance, while the other bootstrap methods provide very conservative intervals.

Within each method, imbalance leads to more conservative intervals. For the parametric and residual methods, coverage is unaffected by this; for the non-parametric case, coverage actually gets better for some reason in most cases. Also, increasing sample size leads to sharper intervals across the board. However the rate of improvement is most pronounced in the case of balanced data, and reduces with imbalance.

Table 4.4: Average lower confidence bound for σ_τ^2 with coverage based on the Thomas-Hultquist and Burdick-Eickman approximations

				Thomas-Hultquist		Burdick-Eickman	
Setting	σ_τ^2	k	b	Coverage	Length	Coverage	Length
1	0.1	10	1	0.980	0.0153	0.980	0.0153
2	0.1	10	2	0.986	0.0121	0.987	0.0104
3	0.1	10	3	0.991	0.0061	0.994	0.0033
4	1	10	1	0.961	0.4369	0.961	0.4369
5	1	10	2	0.970	0.4103	0.970	0.4088
6	1	10	3	0.978	0.3698	0.978	0.3596
7	2	10	1	0.954	0.9490	0.954	0.9490
8	2	10	2	0.967	0.8495	0.967	0.8485
9	2	10	3	0.960	0.8664	0.961	0.8609
10	0.1	20	1	0.988	0.0237	0.988	0.0237
11	0.1	20	2	0.995	0.0237	0.995	0.0224
12	0.1	20	3	0.991	0.0083	0.994	0.0058
13	1	20	1	0.964	0.5593	0.964	0.5593
14	1	20	2	0.973	0.5408	0.973	0.5404
15	1	20	3	0.959	0.5096	0.959	0.5059
16	2	20	1	0.951	1.1488	0.951	1.1488
17	2	20	2	0.972	1.1341	0.972	1.1339
18	2	20	3	0.964	1.1088	0.964	1.1067

For the T-H and B-E methods (Table 4.4), we first notice that the intervals are very conservative when σ_τ^2 is small which is in line with the theory. They progressively improve for large η, and in most cases are better than GCI. Larger sample sizes give sharper intervals with slightly less coverage, but in all cases is over the required 95%. However, the T-H method sometimes estimated a negative bound for low η in which case we set it to 0. This resulted in T-H generally out-performing B-E. We need to go back and check the methodology, and also check whether the same thing happens when we compute a two-sided interval.

5 Discussion

One of the major takeaways from the bootstrap methods is that the non-parametric bootstrap will give the sharpest bounds whereas parametric and residual bootstraps will provide very high coverage. GCI, the Thomas-Hultquist and Burdick-Eickman methods seem to provide a better balance of coverage and sharpness, at least in the one-sided cases.

This opens up a lot of options to try and improve these results. Some form of bias correction can definitely improve our bootstrap intervals. While we can bias-correct directly in the intervals, coverage is generally not an issue other than the non-parametric case. Hence it is also possible to look into bias correction in our point estimates and see how they affect the interval.

While the Thomas-Hultquist and Burdick-Eickman methods provide approximate confidence intervals, Hartung and Knapp proposed an exact confidence interval extending those two ideas. However, it requires some simulations and is part of our next steps. Other approaches that we plan on looking into are Empirical Likelihood, Fiducial Confidence Intervals, as well as Jackknife based approaches. We plan to extend most of these methods for two-sided cases as well. From a wider perspective, we would like to look into what other ways we can enforce 'imbalance' as well. The T-H, B-E and H-K methods all used different configurations and some of them are worth looking into.