Classroom Simulation: Understanding One-Way Random-Effect ANOVA

Eric A. SUESS, Bruce E. TRUMBO, and Yun JIANG, *California State University, East Bay*
Department of Statistics; California State University, East Bay (formerly CSU Hayward); Hayward, CA 94542 USA
eric.suess@csueastbay.edu or bruce.trumbo@csueastbay.edu

Abstract: The one-factor random-effect model is presented. Two simulated datasets are analyzed. and discussed from three points of view: (1) The standard ANOVA table, F test, and method-of-moments estimates of variance components, which lead to a negative estimates of within-batch variance for one of the datasets. (2) Maximum likelihood estimates of variance components. (3) Bayesian probability intervals for such components based on flat priors and computed using a Gibbs sampler. Computations are done in R and WinBUGS.

Key Words: Analysis of variance, random effect model, estimation of variance components, interval estimates, Bayesian estimation, flat (noninformative) prior, Gibbs sampler, teaching, R, WinBUGS.

1. Introduction. A common use of the one-way random-effect analysis of variance model is in a manufacturing situation where a product is made in two stages: first, batches of a precursor are made or selected; second, the final items are produced and measured. The central question is often what relative contributions these two stages make to the observed variability of the final items.

The model. Here we consider the model $Y_{ij} = \mu + A_i + e_{ij}$, where $i = 1, ..., g$ batches, $j = 1, ..., r$ replications within each batch, A_i are independent identically distributed (iid) NORM(0, σ_A), and e_{ij} are iid NORM(0, σ). Thus the variance reflected in the measurement of an individual item is $V(Y_{ij}) = \sigma_A^2 + \sigma^2$. We focus mainly on inferences about the variance components σ_A^2 (among batches) and σ^2 (within batches).

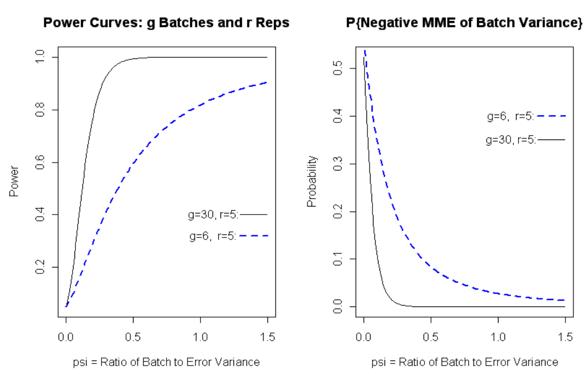

Fig. 1. Power (left) and P{MME(σ_A^2) < 0} for g = 6, 30; r = 5.

2. Traditional ANOVA. The analysis of variance tests H_0: $\sigma_A^2 = 0$ against H_a: $\sigma_A^2 > 0$. The test statistic is $F = MS_B/MS_E$, where $MS_B = SS_B/df_B$, $SS_B = r\Sigma_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$, $df_B = g - 1$, $MS_E = SS_E/df_E$, $SS_E = \Sigma_i \Sigma_j (Y_{ij} - \bar{Y}_{i\cdot})^2$, and $df_E = g(r - 1)$. Under H_0, the statistic $F = MS_B/MS_E$ has the F distribution with df_B and df_E degrees of freedom; we reject H_0 for large values of F. In practice, H_0 is rarely precisely true, so that the salient issue in testing H_0 is by how much $\psi = \sigma_A^2/\sigma^2$ must exceed 0 in order for the F test to have a reasonable chance of detecting that it does exceed 0.

Many textbook examples use small values of g to simplify computation. As shown in the left side of Fig. 1, the power is much better for a larger experiment with $g = 30$ and $r = 5$ (solid line) than for $g = 6$ and $r = 5$.

Method of moments estimates. In real applications, it is often more to the point to estimate variance components than to test H_0. Because $E(MS_E) = \sigma^2$ and $E(MS_B) = r\sigma_A^2 + \sigma^2$, the method-of-moments estimators (MMEs) are MS_E for σ^2 and $(MS_B - MS_E)/r$ for σ_A^2. Unfortunately, the latter estimate takes absurd negative values whenever it happens that $F < 1$. If ψ is close to 0, then such negative estimates are not rare; see Fig 1.

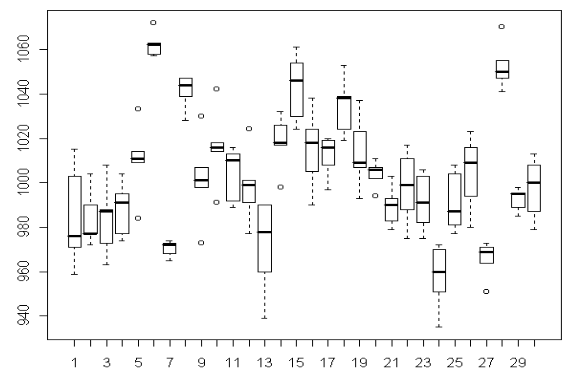

Fig. 2. Boxplots for 30 batches of Dataset 1: $\sigma_A = 25$, $\sigma = 15$.

Source	DF	SS	MS	F	P
Batch	29	91365	3151	15.0	0.000
Error	120	25215	210		

Fig. 3. ANOVA table for Dataset 1: F significant.

Confidence intervals. A confidence interval (CI) for σ^2 is based on the chi-squared distribution with $df_E = g(r-1)$ degrees of freedom. For $df_E = 120$, the ratio of the upper to lower 95% confidence limits for σ^2 is 1.66, computed using

```
qchisq(.975, 120)/qchisq(.025, 120).
```

No exact confidence interval for σ_A^2 is known, and unless g and r are both large enough, the accuracy of approximate intervals may be in doubt. Also, if g is small, any correct CI for σ_A^2 may be too long as to be of practical use.

However, after a little algebra, one can see that an exact 95% confidence interval for ψ has the form $([F/U^* - 1]/r, [F/L^* - 1]/r)$, where F is the F-statistic in the ANOVA table, and L^* and U^* are the 2.5% and 97.5% points, respectively, of the F distribution with degrees of freedom df_B and df_E. Sometimes it is useful to think in terms of the ***interclass correlation coefficient*** (ICC) $\rho_I = \sigma_A^2/(\sigma_A^2 + \sigma^2) = \psi/(\psi + 1)$. The confidence interval $([F - U^*]/[F + (r-1)U^*], [F - L^*]/[F + (r-1)L^*])$ for ρ_I can be derived from the interval for ψ given above. [3, 4]

It is natural to estimate μ as the grand mean of all gr observations, $\bar{Y}_{..} = (1/gr)\Sigma_i \Sigma_j Y_{ij}$, distributed normally with mean μ and variance $(\sigma_2 + r\sigma_A^2)/gr$. But $\sigma^2 + r\sigma_A^2$ is estimated by MS_B, which has $df_B = g - 1$ degrees of freedom. Thus the t distribution can be used to obtain $\bar{Y}_{..} \pm t^*(MS_B/gr)^{1/2}$ as a 95% confidence interval for μ:, where t^* cuts area 2.5% from the upper tail of the t distribution with $g - 1$ degrees of freedom.

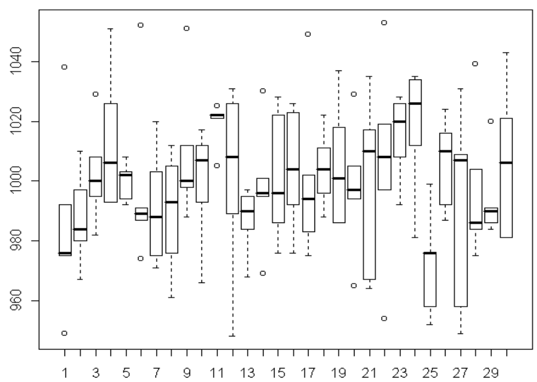

Fig. 4. Boxplots for 30 batches of Dataset 2: $\sigma_A = 1$, $\sigma = 25$.

Source	DF	SS	MS	F	P
Batch	29	14931	515	0.95	0.552
Error	120	65362	545		

Fig. 5. ANOVA table for Dataset 2: F not significant.

3. ANOVA and MMEs for Two simulated datasets. As examples, we use two simulated datasets, each with $g = 30$ and $r = 5$. Thus there are enough batches to get stable results with a variety of methods. Also, we know the true parameter values so we can judge whether our estimates are reasonable. Printouts of these datasets are available online [5].

Dataset 1 has $\mu = 1000$, $\sigma_A = 25$, and $\sigma = 15$. Fig. 2 shows boxplots of the $g = 30$ batches and Fig. 3 shows the ANOVA table with a very highly significant F statistic. The MMEs are 1003 for μ with 95% CI (994.2, 1012.3); 588.1 for σ_A^2 (or 24.25 for σ_A); and 14.5 for σ with CI (12.87, 16.59).

Dataset 2 has $\mu = 1000$, $\sigma_A = 1$, and $\sigma = 25$. Fig. 4 shows boxplots of the batches. The F ratio in the ANOVA table of Fig. 5 is far from significant. Here the MMEs are 1000 for μ with CI (996.4, 1003.7); *negative value* –5.96 for σ_A^2; and 23.34 for σ with CI (20.72, 26.72).

Below we look at methods that give similar results to those above for the "well-behaved" Dataset 1, but also provide more useful results for Dataset 2, which is in some respects analytically "badly-behaved."

3. Maximum likelihood estimators. Observations from different batches are independent. For example, $Y_{11} = \mu + A_1 + e_{11}$ and $Y_{21} = \mu + A_2 + e_{21}$ are independent because A_1, A_2, e_{11} and e_{12} are mutually independent. However, different observations from the *same* batch, say Y_{11} and Y_{12}, are correlated because they have A_1 in common. Two observations Y_{ij} and Y_{ik} from the same batch each have variance $\sigma_A^2 + \sigma^2$, so their covariance is, $\text{Cov}(Y_{ij}, Y_{ik}) = \text{Cov}(A_i + e_{ij}, A_i + e_{ik}) = \text{Cov}(A_i, A_i) = V(A_i) = \sigma_A^2$. Thus their correlation (the ICC) is $\rho_I = \sigma_A^2/(\sigma_A^2 + \sigma^2)$.

Joint density function. Thus the joint density function of the observations Y_{ij} is multivariate normal of dimension br: $f(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{V}) = (2\pi)^{-br/2} |\mathbf{V}|^{1/2} \exp[-(1/2)(\mathbf{y} - \boldsymbol{\mu})'\mathbf{V}(\mathbf{y} - \boldsymbol{\mu})]$, where $\mathbf{y} = (Y_{11}, Y_{12}, ..., Y_{1r}, ..., Y_{br})$ and the mean vector $\boldsymbol{\mu}$ has all gr components equal to μ. The $gr \times gr$ correlation matrix $\mathbf{C} = [1/(\sigma_A^2 + \sigma^2)]\mathbf{V}$ is composed of b^2 submatrices, each $r \times r$. Of these submatrices, $g^2 - g$ have all elements 0. But along the principal diagonal of **C** are b identical submatrices, each with elements 1 along its principal diagonal and elements ρ_I elsewhere.

Computing MLEs. If the density function $f(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{V})$ is viewed as a function of $\boldsymbol{\mu}$ and $\mathbf{V}$ for $\mathbf{y}$ as given by the data, then it is called a likelihood function. The values of μ, σ_A^2, and σ^2, that maximize this likelihood function are called the maximum likelihood estimates (MLEs) of these three parameters of the model. One thing about the MLE for σ_A^2 is clear without computation: it *cannot be negative* because $f(\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{V}) = 0$, except where $\sigma_A^2, \sigma^2 > 0$. Specifically, the MLE of σ_A^2 is $[(1 - 1/g)MS_A - MS_E] / r$, provided this quantity is positive; and the MLE of σ is $\min[MS_E, SS_E + SS_B/gr]$. (See [2].)

Approximate MLEs computed in R are: 1003 for μ with CI (994.3, 1012.2); 23.81 for σ_A with CI (18.14, 31.26); and 14.50 for σ with CI (12.77, 16.45). These results are similar to the ones shown earlier for MMEs, and all intervals easily cover the known true parameter values. (See [5] for the R code.)

MLEs obtained similarly for Dataset 2 are: 1000 for μ, 0.56 for σ_A, (as approximated in R), and 23.13 for σ. However, R does not provide confidence intervals for this dataset because the sample variance-covariance matrix is ill-conditioned. This is not surprising: If we had $\sigma_A = 0$ instead of 1, the dimensionality of the model would have been smaller.

4. The Gibbs sampler using R. The Gibbs sampler is a computational method that involves Markov chains and a Bayesian framework. See [6] for an elementary general introduction to Gibbs sampling and [1] for the details of the

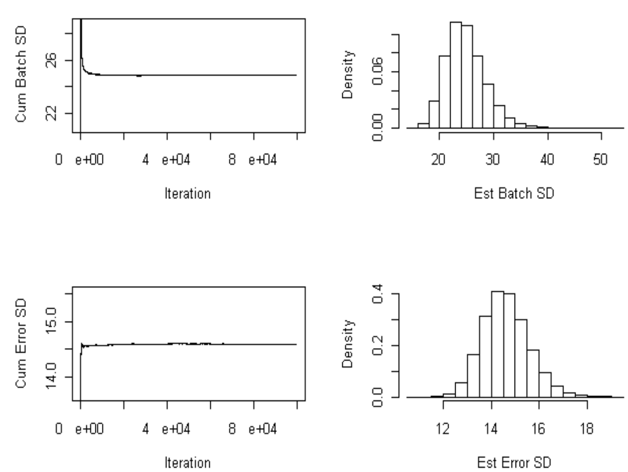

Fig. 8. Graphics for Gibbs Sampler: Dataset 1.

specific situation discussed in this section. Based on prior distributions and the likelihood function, one constructs a Markov chain whose limit, approximated by simulation, provides the corresponding posterior distributions [5].

For Dataset 1, the first column in figure Fig. 8 shows the stability of the convergence to the point estimates of σ_A (upper) and σ, and the second column shows the simulated densities (after burn in: ignoring the first 25,000 of 100,000 iterations) from which the interval estimates are obtained.

For Dataset 2, the simulated densities for σ_A (and thus ρ_I) are strongly right-skewed. Diagnostic graphics show that the Gibbs Sampler is performing properly [5, 6]. Bayesian point and interval estimates obtained from our Gibbs Sampler are shown in Fig. 10. Our Gibbs Sampler used flat priors, so it is not surprising that the resulting Bayesian estimates are in relatively close agreement with the MLEs.

4. Bayesian Results From WinBUGS. WinBUGS [8] is free software widely used in Bayesian analysis. We used two parameterizations to estimate μ, σ_A. σ, and ρ_I. The first parameterization is the same as in Section 3, and the second puts a uniform prior on ρ_I instead of the flat gamma prior on $\tau_A = 1/\sigma_A^2$. WinBUGS results are shown in Fig. 10; those based on the same parameterization as in Section 3 are very similar to the results obtained there [5].

5. Summary Comments. From Fig. 10 we see that all three methods—MME, MLE, and Bayesian—give similar and satisfactory results for Dataset 1, in which σ_A^2 is relatively large. Exact confidence intervals are available for μ, σ, and ρ_I. However, for Dataset 2, in which σ_A^2 is too small to detect with a standard ANOVA, the MME is negative; an ill-conditioned sample variance-covariance matrix prevents R from finding confidence intervals for the ML estimates; while the Bayesian approach of Section 3 yields reasonable point estimates and interval estimates that cover the known parameter values. But in practice, the interval estimates of σ_A for Dataset 2 may too long to be useful. The second parameterization of Section 4 (WinBUGS) gives less-satisfactory interval estimates for σ_A and ρ_I.

REFERENCES

1. Gelfand, Alan E.; and Smith, A.F.M: "Sampling-based approaches to calculating marginal densities," *JASA* **85**, 410 (1990) pp 398-409. [See Sect. 3.4 pp 405-406.]

2. Miller, R.G., *Beyond ANOVA*, Chapman-Hall (1997).

3. Oehlert, Gary W.: *A first course in design and analysis of experiments.* W.H. Freeman (2000), Chapter 11.

4. Snedecor, G.W.; and Cochran, W.G.: *Statistical Methods,* 7th ed. (1980). [Chapter 13 discusses also an alternative model where the intraclass correlation can legitimately be negative.]

5. Suess, Eric A.: `www.sci.csueastbay.edu/~esuess/onewayrand` [Full code for MLE, Gibbs and WinBUGS approaches. Auxiliary instructional materials.]

6. Trumbo, Bruce E.; Suess, Eric A.; and Fraser, C.M.: "Elementary uses of the Gibbs Sampler: Applications to medical screening tests," *STATS*, Winter 2000, pp 3-10. [An elementary exposition of Gibbs samplers.]

7. University of Montana, ST412 website: `www.math.montana.edu/~jarrett/st412/lec17.pdf`. [R code for MLEs]

8. WinBUGS website `www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`. [See Examples for Volume 1, Dyes data; file `www.mrc-bsu.cam.ac.uk/bugs/winbugs/Vol1.pdf`.]

Dataset 1

Method	$\mu = 1000$	"Among" $\sigma_A = 25$	"Within" $\sigma = 15$	ICC = .735
MME (ANOVA)	994 **1003** 1012	**24.25**	12.9 **14.5** 16.6	.610 **.737** .845
MLE (R)	994 **1003** 1012	18.1 **23.8** 31.3	12.8 **14.5** 16.5	**.729**
Bayes (R, Gibbs)	994 **1003** 1012	18.9 **24.9** 33.1	12.9 **14.6** 16.6	.609 **.736** .845
Bayes* (WinBUGS, MCMC)	994 **1003** 1013 994 **1003** 1012	18.8 **24.9** 33.0 18.6 **24.4** 32.1	12.9 **14.6** 16.6 12.9 **14.6** 16.7	.607 **.736** ,844 .600 **.727** .836

Dataset 2

Method	$\mu = 1000$	"Among" $\sigma_A = 1$	"Within" $\sigma = 25$	ICC = .0016
MME (ANOVA)	996 **1000** 1004	**[σ_A^2 = –5.96]**	20.7 **23.3** 26.7	[–.097 **–.011** .137]
MLE ‡ (R)	**1000**	**0.56**	**23.1**	**.0006**
Bayes (R, Gibbs)	996 **1000** 1004	.033 **1.41** 6.63	20.8 **23.3** 26.1	2 e-6 **.009**† .078
Bayes*(WinBUGS, MCMC)	996 **1000** 1004 996 **1000** 1004	.033 **1.28** 6.28 1.01 **5.13** 10.4**	20.8 **23.3** 26.1 20.4 **22.8** 25.7	2 e-6 **.008**† .070 .002 **.07** .17**

* R-Gibbs results and results in the top row for WinBUGS use the same flat priors on μ, σ_A, σ as R-Gibbs; the bottom row of WinBUGS uses uniform prior on ICC, and flat priors on μ and σ.

‡ Because of an ill-conditioned sample variance-covariance matrix, confidence intervals are not available in R.

† Simulated means after burn-in are shown; the corresponding medians are .0007 (Gibbs), .0005 (BUGS).

** An interval estimate that does not cover the known true parameter value.

Fig. 10. Summary of Results. Bold numbers are point estimates (when possible, with 95% intervals in plain type.)