Chapter 2 Variance Components

Paul E. Johnson

February 1, 2016

1 Glossary

Since this book uses unfamiliar notation I'll never remember....

ξ_{ij}: "xi" individual row-level error term for group i row j,

ζ_j: "zeta" a group level random error, j indexes a grouping variable, $Var(\zeta_j) = \psi$

ϵ_{ij}: "epsilon" $Var(\epsilon_{ij}) = \theta$. Individual row-level error uncorrelated with ζ_j

$Cov(y_{ij}, y_{i'j}|\zeta_j) = 0$. Apart from ζ_j, the observed scores are "conditionally uncorrelated"

The combined error term has two variance components

$$\xi_{ij} = \zeta_j + \epsilon_{ij} \tag{1}$$

2 Highlights

1. These are all ANOVA models, in which we aren't primarily interested. Still, need to pick up ideas as they carry over to linear mixed effects.

2. variance component: pieces in (1) above. Could have more components, wish we could disentangle their variances.

3. p. 78: Directed Acyclic Graph representation. Seems unhelpful in this context, but it probably makes psychologists happy.

4. p. 79: Variance of the total error term is the sum of the variances of the components. Because $E[y] = \beta$, $Var(y_{ij}) = E[(y_{ij}-\beta)^2] = E[\xi_{ij}^2] = \ldots = Var(\zeta_j)+Var(\epsilon_{ij}) = \psi+\theta$. This works because we assume the 2 variance components are uncorrelated, so the covariance part $Cov(\zeta_j, \epsilon_{ij}) = 0$

5. p. 79-80. ICC: Fraction of variance that is between grouping units

$$\rho = \frac{\psi}{\psi + \theta} \tag{2}$$

   p. 80. The same number can mean the "within-cluster correlation" of observations. The more tightly inter-dependent are the observations within a cluster, the more distinctive the clusters ought to be.

6. p. 82: Pearson correlation compared to ICC.
7. Hypothesis Tests: Fixed Effects
   a) p. 85. Output includes a "z test" $\hat{\beta}/SE(\hat{\beta})$. Its not a t stat.
      i. It is an "asymptotically valid" test, not precise for a finite sample.
      ii. p. 85. Other software packages offer finite approximations.
   b) Can construct a confidence interval for β in usual way, something like $\hat{\beta} \pm 1.96 std.err.(\hat{\beta})$.
   c) Estimating the standard errors.
      i. "model based" (assuming homogeneous errors within groups, and that across-groups variance is homoskedastic).
      ii. "robust" "sandwich" estimates. (STATA: `vce(robust)`. Disturbingly vague.
         A. Warning: robust behaves badly in small samples.
   d) NOTE to self: work out explanation of the "degrees of freedom" problem and what other software does with finite degrees of freedom. Why is there some "monkey business" going on?
      i. Suppose there are m groups and we fit a dummy variable fixed effects regression. Then do a t-test

$$\frac{\hat{\beta}_1}{std.err(\hat{\beta}_1)} \sim t\, with\, df = N - 2 - m \tag{3}$$

      ii. As there are more groups, the subtraction for df grows more extreme and the required t value for statistical significance grows larger and larger. The fixed effect becomes "less and less" significant.
      iii. We avoid that "loss of degrees of freedom" by estimating a single variance parameter ψ. How many degrees of freedom should be subtracted in order to take that into account? Just 1?
      iv. Here's what is misleading. The random effects model necessarily entails prediction/estimation of the random effects ζ_1, ζ_2,. . ., ζ_m, so in a way, we really are using up degrees of freedom and we should subtract something from the degrees of freedom.
8. p. 88. Hypothesis Tests: Random Effects
   a) Is the true $Var(\zeta_j) = \psi = 0$?
   b) We estimate that with Maximum Likelihood. Then estimate a model without ζ_j. A Likelihood ratio test compares 2 models (LR test). Students should remember this from ML applications like logistic regression.

c) In ML, for a large sample tending to infinity, this number tends to be χ^2 distributed with ν = the difference in number of estimated parameters. (In my mind, the words are "minus two times the log of the ratio of the likelihoods".

$$-2ln(\frac{L_0}{L_1}) = -2\{ln(L_0) - ln(L_1)\} = 2\{ln(L_1) - ln(L_0)\} \tag{4}$$

d) Note 1: any estimate based on data is surely to be bigger than zero, $\hat{\psi} > 0.$, and it could only be exactly 0 if the data was freakishly empty. So the estimated $\hat{\psi}$ is ALWAYS greater than 0. We mean to ask "is it enough greater than 0 to believe that the true ψ is not 0.

e) Note 2: The theoretical complication: the estimator is on the boundary of the sampling distribution. The usual LR test is not exactly correct.

f) p. 89: RHS say that the p value from the LR test is 2 times larger than it ought to be. I have no evidence this is right or wrong. I have heard the argument in other materials that the p value from the LR test is "conservative".

g) p. 91. SCORE test versus LR test versus Wald test. I thought I understood these things until I read this passage. Now I don't understand. Will stick to the LR test, most people do.

h) F test can be used in GLS-based estimator in "xtreg, fe". p. 92: TODO. figure out what Stata output "sigma_u" means. Suspect: standard deviation of intercept estimates in FE model.

9. p. 92-93: DISREGARD or BE CAUTIOUS about Stata output on standard errors and Confidence Intervals output by xtmixed.

10. p. 94: "superpopulation" inference, just as I preached it in class. No such thing as having a data set that includes the "whole population."

11. p. 95. Fixed Effects vs Random Effects.

    a) Target of inference: These particular groups or groups that might be drawn?

    b) Exchangable group random effects: Subjectively, given groups $j = 1, 2, \ldots$, do we believe it could be that they are assigned ζ_j at "random", equally likely, without regard to their index value.

12. p. 99-101. Review assumptions about variance of ζ_j, ϵ_{ij}, the lack of covariance between them.

13. p. 101. Different Estimation methods

    a) ML. p.101. "The idea is to find parameter estimates $\hat{\beta}$, $\hat{\psi}$, and $\hat{\theta}$ that maximize the likelihood function, thus making the responses appear as likely as possible".

    b) Get an intuition from the ANOVA style variance estimates for J groups, n responses (balanced groups). the ML estimators:

i. within cluster variance: Based on squared deviations of individual scores around within-group means. It is "Mean Squared Error", variance of the residuals.

$$\hat{\theta} = \frac{1}{J(n-1)}\left\{\sum_{j=1}^{J}\sum_{i=1}^{n}(y_{ij}-\bar{y}_{.j})^2\right\} \quad (5)$$

$$\frac{1}{J(n-1)}SSE = MSE \quad (6)$$

ii. ML estimate of group random effect. The Model Sum of Squares (MSS) is the squared deviations of group means around the overall "grand mean":

$$\hat{\psi} = \begin{cases} \frac{MSS}{Jn} - \frac{\hat{\theta}}{n} & if\,positive \\ 0 & if\,line\,1\,<0 \end{cases} \quad (7)$$

c) An "unbiased" correction of this from ANOVA = REML

$$\hat{\psi}^M = \frac{MSS}{(J-1)n} - \frac{\hat{\theta}}{n} \quad (8)$$

i. p. 102 "The estimate an be negative, making unbiasedness less attractive than it seems."

d) Unbalanced:

i. p. 102 "Contrary to popular belief, REML is not unbiased for ψ when data are unbalanced. Furthermore, it is not clear which method has the smallest mean squared error (MSE)."

14. p. 109. Predicting (or estimating?) ζ_j.

a) In the way of estimating that RHS consider (NOT the PLS way), the GLS or ML fitting process is geared to give estimates of $\hat{\beta}$, $\hat{\psi}$, $\hat{\theta}$.

b) From those, we formulate a statement about the values of ζ_j that are most likely to have produced y_{ij}.

c) EMPIRICAL Bayes. Prior(ζ_j) is $Normal(0,\hat{\psi})$. Result. The Empirical Bayes estimate is a "shrunken" version of the ML estimate.

$$\hat{\zeta}_j^{EB} = \frac{\hat{\psi}}{\hat{\psi}+\hat{\theta}/n_j}\hat{\zeta}_j^{ML} \quad (9)$$

d) The shrinkage coefficient, $\hat{R}_j = \frac{\hat{\psi}}{\hat{\psi}+\hat{\theta}/n_j}$ is similar to the ICC, but the individual error is down weighted because it is divided by the number of observations within group j .

e) If a group has LOTS of observations, not much shrinkage is applied. But, when FEW observations, then much shrinkage is applied.

15. p. 113. Empirical Bayes standard errors. I am pretty sure all of this applies only to Variance Components models, not necessarily to regression models. I'll dog ear this, maybe come back later.

   a) Comparative standard errors. Think this through.

      i. Suppose a score y "pops out" from nature. After that, we can triangulate on ζ_j, say some values are more likely than others.

$$Var(\zeta_j|y_{1j}, y_{2j}) = \frac{\theta/n_j}{\psi + \theta/n_j}\psi = (1 - R_j)\psi \tag{10}$$

      ii. CLAIM p. 113. The variance of ζ_j "is also the conditional variance of the prediction errors $\hat{\zeta}_j^{EB} - \zeta_j$,given the observed y. The variance of our prediction of ζ_j given the observed data, $Var(\hat{\zeta}_j^{EB} - \zeta_j|y_{1j}, y_{2j})$.

      iii. Mean Squared Error of Prediction. See Stata below: RSES=random effect standard errors

   b) Diagnostic Standard Errors: spot outliers in ζ_j

      i. The variance of EB predictions (over repeated samples)

$$Var(\hat{\zeta}_j^{EB}) = \frac{\psi}{\psi + \theta/n_j}\psi \tag{11}$$

      ii. That's the same shrinkage factor

      iii. p. 114: used to spot "outliers". If a ζ_j is outside 2 standard deviations, it is in the extreme zone.

3 Stata notes

1. xtreg. Requires xtset first to specify grouping. Less well integrated with new stata tools.

   a) FGLS: Feasible generalized least squares is an option for xtreg, re.

   b) Requires `xtset id` , so following xtreg knows the grouping indicator on rows

2. xtmixed. Grouping set as part of the command.

   a) reml mle option: ML estimate

   b) reml option: REML

   c) Notation: `xtmixed y x || id:   , mle` fits y with predictor x and random effect with variable named "id". The empty space after colon means no slopes depend on "id".

3. p. 83: Data Preparation

```
reshape long wp wm, i(id) j(occasion)
```

a) p. 83. Note Stata uses i and j BACKWARDS from this book.

b) p. 85. In time series tools, Stata uses t for rows within group i.

c) Stata uses u to refer to the grouped random effects in the model, and e for the individual row random errors.

d) p. 85. Note ICC is not included in xtreg output, user can calculate from random effects.

4. p 112. POST estimation extraction of estimated values

   a) `_b[varname]` retrieves an estimated fixed effect

   b) To estimate standard deviations, fit differently

      `xtmixed wm || id:  , mle estmetric`

      i. Extract $ln(\sigma)$ like this

      p. 112. "The estimation metric for each standard deviation is the logarithm of the standard deviation. We can access the estimated logarithms by using the syntax `[lns1_1_1]_cons` and `[lnsig_e]_cons`."

      i. TODO: Find out how to make a Stata session tell me all of the things "out there in memory" that I can retrieve.

   c) Use the predict with reffects to pull out guesses about random effects, group-by-group

      pp. 112 shows "manually calculated" EB and compares

      results from Stata predict method. Same idea as lme4 "ranef"

```
predict eb2, reffects
sort id
format eb1 eb2 %8.2f
list id eb1 eb2 if occasion==1, clean noobs
```

   d) p. 114. Stata `predict_se, rses`. RSES=random effect standard errors