

Multilevel Models

2. Random Intercept Models

Germán Rodríguez

Princeton University

March 28, 2018

Random intercepts

We now consider models with covariates, starting with the random-intercept model

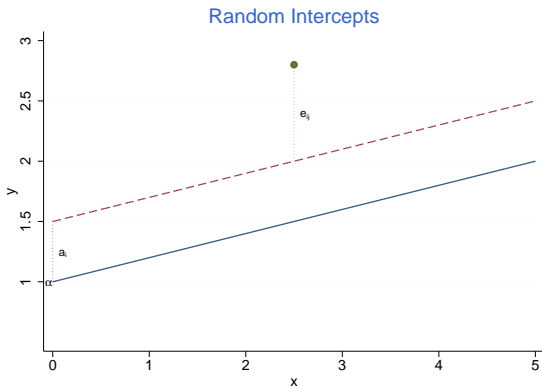
$$Y_{ij} = \alpha + a_i + x'_{ij}\beta + e_{ij}$$

where Y_{ij} is the outcome for the j -th individual in the i -th group, α is the constant, and x_{ij} is a vector of predictors with coefficients β .

We have two residuals: a group random effect $a_i \sim N(0, \sigma_a^2)$ and an individual effect $e_{ij} \sim N(0, \sigma_e^2)$, assumed independent of each other and of the covariates.

Given the random effect a_i , the outcome $Y_{ij}|a_i$ follows an ordinary linear model with intercept $\alpha + a_i$, hence the name “random intercept”.

Here's the model in graphical form



In terms of our example, we assume that language scores depend on verbal IQ with a common slope and an intercept that varies across schools.

Unconditional moments

In this model the expected value of the outcome is

$$E(Y_{ij}) = \alpha + x'_{ij}\beta$$

The variances and covariances of the outcomes are

$$\text{var}(Y_{ij}) = \sigma_a^2 + \sigma_e^2 \quad \text{and} \quad \text{cov}(Y_{ij}, Y_{ik}) = \sigma_a^2, j \neq k$$

The correlation between any two outcomes in a group, or *intra*class correlation is

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

These results are exactly the same as in the variance-components model, the only difference is that we now account for covariates.

Estimation of the parameters

The OLS estimator of α and β ignoring the correlation structure is consistent but not fully efficient. The associated standard errors need to be corrected for clustering.

A better approach is to use maximum likelihood (ML). This is implemented in Stata's `xtreg`, `mle` and `mixed`, `mle`, as well as R's `lmer()` if you specify `REML = FALSE`.

Alternatively, one can use restricted maximum likelihood (REML), as implemented in Stata's `mixed`, `reml`, or as the default in R's `lmer()`, which relies on error contrasts to estimate the variance components.

Given estimates of σ_a^2 and σ_e^2 , both ML and REML estimate α and β using generalized least squares. The two procedures give very similar estimates if the number of groups is large.

The computing logs fit this model to the language score data as a function of verbal IQ centered on the grand mean, so the model is

$$Y_{ij} = \alpha + a_i + \beta(x - \bar{x}_i) + e_{ij}.$$

The fitted equation is

$$E(Y_{ij}) = 40.609 + 2.488(x - \bar{x})$$

We estimate the variances as

$$\hat{\sigma}_a^2 = 3.082^2 \quad \text{and} \quad \hat{\sigma}_e^2 = 6.498^2$$

The intraclass correlation is

$$\hat{\rho} = \frac{3.082^2}{3.082^2 + 6.498^2} = 0.167,$$

so schools account for 17% of the variation in language scores after taking into account verbal IQ.

In ordinary linear models we compute R^2 for model ω as the proportionate reduction in the RSS starting from the null model ϕ

$$R^2 = 1 - \frac{\text{RSS}(\omega)}{\text{RSS}(\phi)}$$

In a two-level random intercept model we can define R^2 as

$$R^2 = 1 - \frac{\hat{\sigma}_a^2(\omega) + \hat{\sigma}_e^2(\omega)}{\hat{\sigma}_a^2(\phi) + \hat{\sigma}_e^2(\phi)} = 0.384$$

This statistic can also be calculated by level

$$R_a^2 = 1 - \frac{\hat{\sigma}_a^2(\omega)}{\hat{\sigma}_a^2(\phi)} = 0.511 \quad \text{and} \quad R_e^2 = 1 - \frac{\hat{\sigma}_e^2(\omega)}{\hat{\sigma}_e^2(\phi)} = 0.346$$

Unlike linear models R^2 is not guaranteed to increase when variables are added!

Predicting random effects

Consider now estimating the group intercepts $\alpha + a_i$, which involves predicting a_i given the other parameters in the model

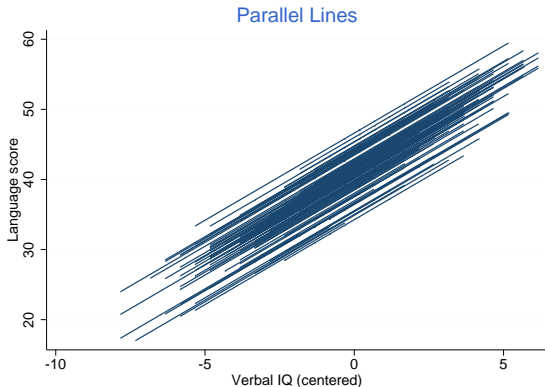
The ML estimator of a_i is obtained by treating everything else as known and maximizing the likelihood, and turns out to be the group means of the residuals $y_{ij} - (\hat{\alpha} + x'_{ij}\hat{\beta})$.

The EB estimator maximizes the posterior distribution, obtained as the product of the likelihood and prior, and can be obtained in Stata using `predict, reffects` after `mixed` (but not after `xtreg`), and in R via `ranef()`.

Comparing the EB and ML estimators you should expect the usual shrinkage towards zero, as we'll soon see.

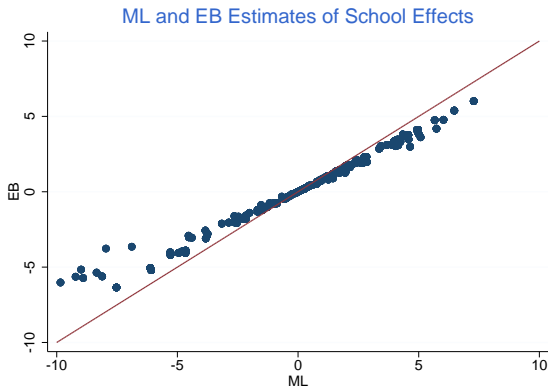
Parallel lines

In the computing logs we fit the model using maximum likelihood and then obtain the fitted values $\hat{y}_{ij} = \hat{\alpha} + \hat{\alpha}_i + x'_{ij}\hat{\beta}$. The figure below shows these lines.



Comparison of ML and EB

We can also compare the empirical Bayes estimates with the maximum likelihood estimates of the school effects.



The computing logs have a similar plot using estimated intercepts rather than school effects.

To test hypotheses about β , for example the hypothesis $H_0 : \beta_2 = 0$ for a subset β_2 , we can use

- 1 Wald tests, constructing the quadratic form $W = \hat{\beta}_2' \text{var}(\hat{\beta}_2)^{-1} \hat{\beta}_2$, which is asymptotically χ_p^2 with d.f. equal to the number of coefficients in β_2 .
- 2 Likelihood ratio tests, where we fit the model with and without the predictors involved in β_2 and take twice the difference in log-likelihoods, which is asymptotically χ_p^2 with the same d.f. as above.

With ML both tests are available. With REML we can only use Wald tests; because the models with and without β_2 use different error contrasts the restricted likelihoods are not nested!

Tests about the variance components proceed as before.

For the language scores data the estimated slope of 2.49 has a standard error of 0.07 leading to a Wald $z=35.5$ (equivalent to $\chi_1^2 = 1261.4$) and a 95% confidence interval of (2.35, 2.63).

The likelihood ratio test compares the log-likelihoods with and without verbal IQ and gives $\chi_1^2 = 1001.5$. In multilevel linear models the LR and Wald tests are \neq , but asymptotically equivalent.

The LR test is not available if you use REML, which used to be the default of `mixed`. If you try, Stata will warn “REML criterion is not comparable under different fixed-effects specifications”

Tests for the variance components are as before. In our example the test for σ_a^2 is $\bar{\chi}_{01} = 225.92$ using ML and $\bar{\chi}_{01} = 227.30$ using REML. School effects are clearly significant.

Between-groups estimation

Consider the group means, $\bar{Y}_i = \sum_j Y_{ij}/n_i$, which follow the model

$$\bar{Y}_i = \alpha + a_i + \bar{X}_i' \beta + \bar{e}_i$$

where \bar{X}_i is the average of the covariates and $\bar{e}_i \sim N(0, \sigma_e^2/n_i)$ is the average error term.

These means are independent and we can estimate α and β by OLS or WLS.

Stata can compute this estimator via the command `xtreg, be`. The option `wls` uses group sizes as weights. (Ideally, of course, we would like to use weights inversely proportional to the variances of the group means.)

The between-groups estimator of the slope for the language score data using WLS is 3.90, much larger than the RE estimate of 2.49.

Within-groups estimation

Consider now the differences between the individual outcomes and the group means. These follow the model

$$Y_{ij} - \bar{Y}_i = (X_{ij} - \bar{X}_i)' \beta + (e_{ij} - \bar{e}_i)$$

Note that α drops out, as would any variables which are constant within groups. (None in our example.)

The estimator based on within-group variation is known as the **fixed effects** estimator, and is equivalent to using a dummy variable for each group. It is available in Stata in `xtreg`, `fe`, and in R using the package `plm`.

The within-groups estimate of the IQ slope for the language score data is 2.41.

Between and within-groups together

The random effects estimator is a weighted average of the between and within estimators.

It is possible to obtain the within and between-groups estimates together by fitting a model that includes as predictors the group means and the differences from the group means:

$$y_{ij} = \alpha + a_i + \bar{x}_i\beta_B + (x_{ij} - \bar{x}_i)\beta_W + e_{ij}$$

For the language score data we obtain estimates of

$$\hat{\beta}_B = 4.00 \quad \text{and} \quad \hat{\beta}_W = 2.41$$

The between estimator differs slightly from the WLS estimator because the weights are not the exactly the same. The within estimator is identical to the fixed-effects estimator.

A Wald test of equality gives $\chi_1^2 = 25.79$ and casts doubt on the validity of the random effects estimator.

The Hausman specification test

Hausman has proposed a specification test for the random effects model and the assumption that the school effects are exogenous

If the group effects are in fact independent of observed covariates then the random effects estimator is both consistent and efficient.

If the group effects are correlated with observed covariates, then the fixed effects estimator is consistent but not efficient.

The Hausman test is based on the quadratic form

$$(\hat{\beta}_E - \hat{\beta}_C)' [\hat{\text{var}}(\hat{\beta}_C) - \hat{\text{var}}(\hat{\beta}_E)]^{-1} (\hat{\beta}_E - \hat{\beta}_C)$$

where I used E for efficient (here the random-effects estimator) and C for consistent (here the fixed-effects).

For the language score data the Hausman test gives $\chi_1^2 = 33.75$, strong evidence of model misspecification.

Fixed or random?

A random-effects model that fails the Hausman test is often abandoned in favor of the fixed-effects model. A few caveats:

- If the test rules out RE it doesn't mean FE is the correct model! Both models assume uncorrelated errors at level 1. It is only omitted variables at level 2 that are handled by FE.
- Using FE precludes estimating coefficients for variables that are constant within a group (for example school SES). Sometimes these effects are of primary interest.
- An alternative approach is to include the group means as predictors for any variable where the between and within group estimators are significantly different.

Centering on group means

Centering the individual predictors on the group means is *optional* when the group means are included in the model.

To see this point write

$$\bar{x}_i' \beta_B + (x_{ij} - \bar{x}_i)' \beta_W = \bar{x}_i' (\beta_B - \beta_W) + x_{ij}' \beta_W$$

So all that happens if x_{ij} is not centered is that the coefficient of \bar{x}_i becomes the *difference* between the between and within coefficients, which is convenient for testing.

Centering the individual predictors on the group means does not make a lot of sense if the group means are not included in the model. Centering on the grand mean is fine in all models that include a constant.

What happens if the random-intercept model is correct but we ignore the clustering?

- As noted earlier, the OLS estimate of the fixed effects is consistent but not fully efficient, so this is not a serious problem in large samples
- The estimated standard errors, however, are incorrect. A common misconception is that they are always too small. As shown in MALMUS §3.10
 - They are too small for between-cluster covariates
 - but too large for within-cluster covariates

The solution is to adjust for clustering or, better still, use ML.

Exercise. Verify the last statement using the language scores in the website, fitting a model with verbal IQ and school SES as examples of within and between predictors using OLS and ML.