Applied Multilevel Regression Modeling

Day 6: Generalized Linear Mixed Models

Constantin Manuel Bosancianu

WZB Berlin Social Science Center
Institutions and Political Inequality
bosancianu@icloud.com

August 5, 2019

Welcome to week 2!

Last week

Multilevel models are a great way of systematically explaining group-based heterogeneity, rather than treating it as an estimation nuisance.

Important quantities: fixed effects (coefficients) and random effects.

Model fit: likelihood-based indicators (*logLikelihood*, deviance, AIC, BIC), mostly used in relative comparisons.

Cross-level interactions: a way of explaining cross-context variation in an effect.[1]

[1] Sometimes this won't be possible due to low sample size—remember, J is the sample size that matters for L2 parameters. In this case, it's more honest to admit to the limitations. See Andrew Gelman's suggestion of the *secret weapon* strategy, when MLM is not feasible: `https://statmodeling.stat.columbia.edu/2005/03/07/the_secret_weap/`.

Notation—pays off

$$\begin{cases} EFF_{ij} = \beta_{0j} + \beta_{1j} * EDU_{ij} + \beta_{2j} * INC_{ij} + \epsilon_{ij} \\ \beta_{0j} = \gamma_{00} + \gamma_{01} * CORR_j + \gamma_{02} * INEQ_j + \upsilon_{0j} \\ \beta_{1j} = \gamma_{10} + \gamma_{11} * CORR_j + \upsilon_{1j} \\ \beta_{2j} = \gamma_{20} + \gamma_{21} * INEQ_j + \upsilon_{2j} \end{cases} \tag{1}$$

Estimated parameters are in red.

Keep in mind that by default `R` will also estimate $\rho\sigma_{\upsilon_{0j}}\sigma_{\upsilon_{1j}}$, $\rho\sigma_{\upsilon_{0j}}\sigma_{\upsilon_{2j}}$, $\rho\sigma_{\upsilon_{1j}}\sigma_{\upsilon_{2j}}$.

This week

Extensions of the standard MLM format—two-levels, hierarchical, continuous outcome:

- categorical dependent variables: GLMMs;
- observations over time: growth-curve modeling in MLM framework;
- non-hierarchical nesting: cross-classification and multiple-membership;
- post-stratification for sub-group estimates: MrP;
- spatial dependence among errors: spatial MLM.

Each are case studies in how to adapt the MLM framework to particular research settings.

The framework is even more flexible than this: Poisson, survival analysis, IRT, dyadic analysis!

GLMs as a starting point

Why not OLS (1)

Imagine (latent) process Y, which takes observed values 1 or 0 with probability π_i and $1 - \pi_i$: turnout, outbreak of civil conflict, being asked for a bribe, bankruptcy etc.

This produces a Bernoulli distribution[2] with parameter π.

For such a distribution, the mean is π, while the variance is $\pi(1 - \pi) \Rightarrow$ not independent of the mean.

The assumption of homoskedasticity will not be met in this case.

[2] A special case of the binomial distribution.

Why not OLS (2)

Our model would try to estimate π, which is bounded between 0 and 1.

There is no way to control predictor values $\Rightarrow$ predictions can frequently be out-of-bounds (< 0 or > 1).

Two solutions to this problem, through the use of a "translator function":

- ✓ convert π from $(0; 1)$ to $(-\infty; \infty)$: logit function
- ✓ convert $X\beta$ from $(-\infty; \infty)$ to $(0; 1)$: inverse logit function

The case of turnout

This "translator" function is called a *link function* (easy to remember, because it links the probabilities to the linear regression framework).[3]

Take the case of turnout:

- ✓ There is an underlying probability, π, which determines whether a person shows up at the ballot box or not
- ✓ The dependent variable (0 or 1) has a standard binomial distribution, $\mathcal{B}(\pi, n)$
- ✓ We have to predict π using our IVs: $X1$, $X2$ ...

[3]The *logit* is the canonical link function for the Bernoulli distribution, but one could also work with the *probit* link function.

The *logit*

$$logit(\pi) = log(\frac{\pi}{1-\pi}) \tag{2}$$

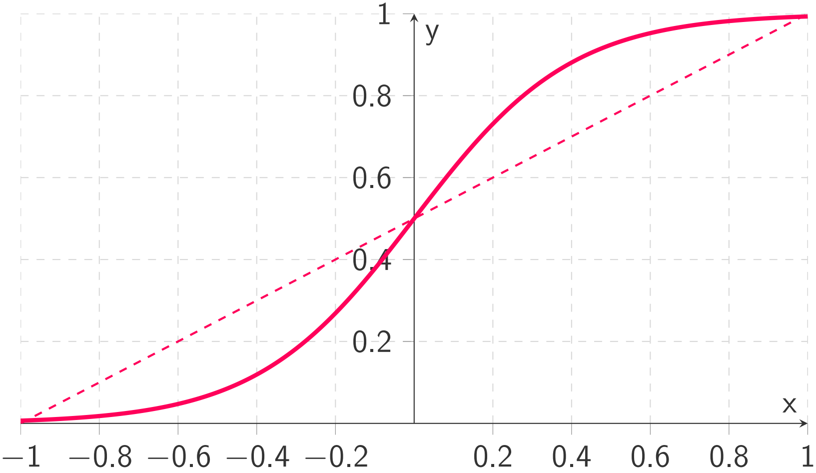
It's pretty simple to visualize that $\lim_{\pi \to 1} \frac{\pi}{1-\pi} = \infty$ and that $\lim_{\pi \to 0} \frac{\pi}{1-\pi} = 0$.[4]

The logarithm simply maps this to the entire $(-\infty; \infty)$ space. Any quantity in $(0; 1)$ gets mapped by the logarithm to the $(-\infty; 0)$ space.

$$logit^{-1}(x) = \frac{e^x}{1+e^x}, \text{does the opposite conversion} \tag{3}$$

[4]The fractions are called "odds".

Non-linear relationship



Coefficient interpretation

$$log_e(\frac{p}{1-p}) = \beta_0 + \beta_1 X1_i + \cdots + \beta_k Xk_i \quad (4)$$

We pay a "price" for this "translation": the βs are expressed in log-odds units.

A simple exponentiation produces odds ratios, which are *slightly* more intuitive:

$$OR = \frac{\frac{P(Y=1|x=1)}{P(Y=0|x=1)}}{\frac{P(Y=1|x=0)}{P(Y=0|x=0)}} \quad (5)$$

ORs range between 0 and ∞, with an OR=1 indicating no effect.

Prob.–Odds–Log odds

Probability p	Odds $\frac{p}{1-p}$	Log odds (logit) $log_e(\frac{p}{1-p})$
0.01	1/99=0.0101	-4.60
0.05	5/95=0.0526	-2.94
0.25	25/75=0.3333	-1.10
0.50	50/50=1	0
0.75	75/25=3	1.10
0.95	95/5=19	2.94
0.99	99/1=99	4.60

Adapted from Fox (2016, p. 377)

No error?

It exists, but it's difficult to formalize.

For a set of predictors that determine a mean π, there are 2 errors:

- ✓ $1 - \pi$, with probability π
- ✓ $0 - \pi$, with probability $1 - \pi$

For other predictor values that determine a mean π', the errors are different, which means no common error distribution exists.

Multilevel logistic

Dichotomous response

The model set-up is similar to that of a linear MLM, with the exception that the L1 model is a logistic one.

Below you have the simplest case: the null model.

$$\begin{cases} log_e(\frac{p}{1-p}) = \beta_{0j} \\ \beta_{0j} = \gamma_{00} + \upsilon_{0j} \end{cases} \tag{6}$$

$$log_e(\frac{p}{1-p}) = \gamma_{00} + \upsilon_{0j} \tag{7}$$

ICC

With only υ_{0j} there is no clear way of estimating the ICC.

A useful shortcut is to consider the variance of the e_{ij} as fixed at $\frac{\pi^2}{3}$ (the variance of the standard logit distribution) (Merlo et al., 2006, pp. 291–292).[5]

In this case, an approximation of the ICC is $\frac{\sigma^2_{\upsilon_{0j}}}{\sigma^2_{\upsilon_{0j}} + \frac{\pi^2}{3}}$.

[5] That's the actual π (3.141592...), not the symbol for probability we've been using.

Interpretation

Although it's a mixed-effects specification, the interpretation of the coefficients is similar to that in a single-level logistic regression: log-odds.

For an easier interpretation they can either be transformed back in odds (e^{β}), or directly in predicted probability change in the outcome.

The software will typically compute effects for the "average case".

"Divide by 4 rule": logistic curve is steepest at the midpoint, where the slope is $\beta\frac{e^0}{(1+e^0)^2} = \frac{\beta}{4}$.

Estimation

Combining mixed-effects models with GLM results in a very complex estimation procedure.

- ✓ Taylor series expansion: a secondary iterative procedure (an approximation of the nonlinear link function).
- ✓ Numerical integration through approximate calculations (with a sufficient number of quadrature points).

Both of these have serious computational requirements.

References

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., . . . Larsen, K. S. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, *60*(4), 290–297.