Multilevel analysis

Course summary

Emmeke Aarts (E.Aarts@uu.nl)
Methodology and Statistics
Faculty of Social and Behavioral Sciences

We are using MLM when we have nested data which means there're some of natural clustering going around in the data.

When we have nested data, it's good to take this nestedness into account. otherwise we see inflation or bias in p-value. they'll become too low. But also in ML analysis, it allows us to use the variables at the right measurement level.

The Multilevel Regression Model

1st level

At the lowest (individual) level we have

- $y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij}$

person i in cluster j = cluster specific intercept β_{0j} + cluster-specific slope β_{1j} x predictor X_{ij} + residual error = diff. between what we predict and what we expect from person i in cluster j

2nd level

and at the second (group) level

- $\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$
- $\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$

u terms: cluster specific deviations are not returned in the output. Instead, we assume they follow $\sim N(0, \sigma^2)$ (σ^2_{u0}: intercept variance / σ^2_{u1}: slope variance) this tells sth. about how much variation between clusters either in intercepts or slopes

(any remaining unexplained differences're captured in u terms)

combining (substitution and rearranging terms) gives

- $y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}}_{\text{fixed effect (this is just normal reg)}} + \underbrace{u_{0j} + u_{1j}X_{ij} + e_{ij}}_{\text{random effect (variance terms)}}$

" They're unconnected from e_{ij}, but they're connected to each other ⇒ Multivariate normal dist

→ MLM is just a linear reg. w/ complicated error terms distributed over multiple levels.

*also have covariance term ** Don't forget in Exam

β_{0j} cluster-specific intercept is composed of: overall average γ_{00} (overall intercept based on all the observations) + $\gamma_{01} Z_j$ + u_{0j}

Z_j: predictor at 2nd level partly explains why the intercepts differ over clusters

u_{0j}: cluster specific deviation to the overall intercept

clusterspecific

β_{1j}: same rationale holds for the slopes.

composed of: overall slope γ_{10} (overall effect we observe over all our clusters) + $\gamma_{11} Z_j$ + u_{1j}

$\gamma_{11} Z_j$: we can also try to explain it by level 2

u_{1j}: cluster specific deviation to this overall slope

* Step 1) intercept-only model: no predictors.
only predicting the overall mean over clusters
& decompose the variance → variation within clusters
& between clusters: when intercept variance (level 2 var) is significant, this tells us if we need ML analysis at all.

predictors.
possible crosslevel interaction that explains why the slopes differ over clusters.

3/10/22

Steps in the analysis

Step 1: Intercept only (i.e. no explanatory variables)

- Test whether level 2 variance is significant : Use dev. diff test, cuz Wald Test assumes certain dist. which is a mismatch between what's actually going on
  - Wald test: one sided test for variances so divide reported p-value by 2
  - Better test: deviance difference with model *without* a variance at level 2 (model 0 in many of the Excel files)

  But, whatever you do, divide p-val/2, cuz the variance cannot be neg. therefore, it's always one-sided test!
- Calculate ICC value : to see the amount of nesting / dependencies within our data. ICC ranges from 0 to 1 ~ gives the idea of how strong the nesting actually is within our data set.
- Use as baseline for calculating R^2 values
  - except for longitudinal data !

3

3

Steps in the analysis

Step 2: add level 1 predictors

- Eliminate predictors that have non-significant effects
- Calculate proportion explained variance at levels 1 and 2 using our baseline model (intercept-only model)

Step 3: add level 2 predictors

- Eliminate predictors that have non-significant effects
- Calculate proportion explained variance at level 2 only becuz the variable only differs over clusters, and not within cluster. So by default, it cannot explain any differences within the cluster

4

4

1) For ex, does the effect of Extra. on Popularity differ between schools?
does " of IQ on reading scores " classes?

2) → Again, we're testing a variance component → divide P-val /2
> Even when the fixed effect of lv.1 predictors are not sig, we still test for slope variance!!
becuz it could be that there're actually a lot of variation going on such that some are positive & some are neg. across diff. clusters, and if we average them over all the clusters, we get a flat effect. So it could be actually very interesting... that it varies so much so that the sign is opposite in clusters ⇒ So re-introduce the non-sig. level1 predictors into the model & test for slope variance!

3/10/22

3) If you have 2 random slopes, that means you have covariance between intercept & randomslope1, int. & rs2, and
5) predictive intervals give us an intuitive interpretation on what a slope variance covariance between rs1 & rs2!! → becomes this matrix actually means, & if it's large or small. Ex) 95% of the slopes over clusters between -2 and 12
↳ gives a very intuitive interpretation what slope variance actually means for overall effect that we see, especially when it turns out the signs are opposite for some of the clusters

Steps in the analysis

1 Step 4: add random slopes ~ checking if the effect of level 1 predictor varies across the clusters or not!

- Test for variance: Again divide reported p-value by 2 cuz variance component cannot be negative
- 2 Even when the fixed effect of a level-1 predictor variable is not significant in step 2, the random effect of this variable may be significant
  - also check variables that were eliminated in Step 2!
- Add random effects one by one
- 3 Also report the covariance between random intercept and random slope
- 4 Do not calculate proportions explained variance for intercept and error term becuz we're just decomposing the variance in a diff. way & we're not explaining any of the variance.
- 5 Calculate* predictive intervals

6 → If we have sig. random slopes, to interpret these to see what they actually mean, we calculate the predictive intervals! to get a better feeling of the magnitude of the variation in slopes over the clusters.

6) that was the whole goal: we wanted to explain why the slopes differ over the clusters.
So we're also calculating how much we explain here.

Steps in the analysis

Step 5: Cross-level interaction ~ If we had sig. slope variance, we can see if we can partly explain why this effect differs over the clusters using variables at level2.
So we only do this step when we in fact have sig. random slope.

- 1 Only for significant random slopes
- 2 Predict with variables at level 2 : so we make this crosslevel interaction between the var. that has sig rs and the level 2 predictor.
- 3 Also include level-2 variables that were not significant at step 3
- 4 Always, if interaction is included, also include main effects
- 5 Interpret interaction effect and main effects simultaneously ~ they should be interpreted as a system w/ 2 main effects & an interaction.
- 6 Calculate proportion explained variance for random slope

hopefully explain the slope variance partly... why the effect differs over the clusters.
To see how well this crosslevel interaction helps you w/ this, we calculate R^2 for random slope.

3) Even if they're not able to predict the overall outcome, they might shed info. on why the effect itself differs over the clusters. Don't forget to remove 'em again if the interaction turns out to be not sig.
4) Also don't forget to include the main effect if you're using them for interaction Term.

5) You should always interpret this as a system if there's an interaction effect, becuz they're all connected to each other : 2 main effects & interaction effect, the interpretation of one main effect is now dependent on the value of the other main effect ~ "I wanna see the interpretation of all of 'em together as a system"

Exceptions I: Longitudinal data analysis

- Baseline model for calculating R^2 is model with time, *not* the intercept only model !!
  - Ignoring time results in an overestimate of the variance component at the lower level & underestimate the variance at the upper level
    ⇒ giving you a wrong benchmark model to calculate R^2
- Intercept only model still used for calculating ICC, though !!
  becuz if you use the model w/ Time, your ICC represents the amount of variation at the 1st measurement occasion
  & if you don't include Time, it represents the amount of variation over all the occasions.
  → So the interpretation slightly differs.

7

2.) Becuz the level 1 variance doesn't differ over the models becuz it's always fixed, it's tricky to calculate R^2.
In order to do that still, instead of focusing on an unexplained variance at the lower level,
we found a way to calculate the explained variance of fixed predictors in our model: σ^2_F represents explained variance of the linear predictors. We use this to compute both Overall Explained variance & Unexplained variance at level 1 and level 2
3) Also becuz the lower level variance is fixed, the model is being re-scaled every time,
So that $\sigma^2_R = 3.29$. Therefore, we cannot compare the reg. coef. over diff. models cuz it rescales all the time.

Exceptions II: Binary logistic regression

1. Step 1: variance on level 1 is not estimated becuz it's fixed! cuz it's a latent construct we need to bound it
   - Use $\sigma_R^2 = \pi^2/3$ = 3.29 for level 1 variance in calculation of ICC
2. Step 2: calculate variance σ_F^2 of fixed part of the linear predictor and use to calculate (un)explained variance
3. Be aware estimates change because of scaling the underlying latent variable
4. Interpretation of parameters:
   - Qualitative: if X increases then probability increases/decreases : only looking at the sign of reg. coefficients.
   - Quantitative: if X increases one unit then the logit increases with ... units
   - Quantitative: if X increases one unit then the odds become ..."times" greater
     ↳ take exponent of reg. coef. → odds : multiplicative !!

8

8

So now we have not 2 outcome categories, but more than 2 & they're ordered.

Exceptions III: Ordinal logistic regression

- "Cumulative" logit models, so multiple logistic regressions
  - Category 1 vs the rest, 1 & 2 vs the rest, etc
  - Assumes proportional odds : the effect of predictors is equal over the diff. multiple logistic regressions we fit.
- Thresholds only important when predictions are made : predicting prob. of being in that category.
  - i.e. when (marginal) probabilities calculated
- Variance calculations σ_R^2 and σ_F^2 as for binary logistic regression
- Interpretation of parameters:
  - Qualitative and quantitative as for binary logistic regression

same as in binary logistic regression.

Meaning that the value of the reg. coef. is fixed over the diff. logistic models we're fitting. The only thing that differs is the threshold value, which signals us when do we switch from one outcome category —to→ the other!

9

9

100%

Predictive intervals vs confidence intervals

FIXED	estimate	S.E.
mean/intercept	0.74	0.23
sex (ref=boy)	1.25	0.03
extrav	0.45	0.02
text	0.09	0.01
text with extrav		
RANDOM	**estimate**	**S.E.**
VAR(e(ij))	0.55	0.02
VAR(u(0j))	1.28	0.31
VAR(u(extr))	0.03	0.01
Covar(u(0j),u(extr))	-0.19	0.05

popularity data

We have rs for Extra meaning that the effect of Extra on popularity differs over the classes

10

So, the overall effect of Extra w/ 1 point increase in Extra, the popularity is predicted to increase by 0.45. (given that texp is 0, if there were to be an interaction Extra * texp)

Slope var. of Extra = 0.03, which seems to be quite small... but let's see what happens if we calculate predictive intervals!

Predictive intervals vs confidence intervals

- 95% confidence interval for mean effect of extraversion
  - $0.45 \pm 1.96 * 0.02$=[0.41; 0.50] → meaning that if we repeatedly take samples, & calculate the 95% conf. intervals then 95% of intervals would contain the population mean
- If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain the population mean.

≠ interpretation differs.

- 95% predictive interval for effect of extraversion
  - $0.45 \pm 1.96 * \sqrt{0.03} = [0.09;\ 0.81]$ → meaning that 95% of the reg. coefficients over the diff. clusters for extraversion are between 0.09 and 0.81.
- 95% of the regression coefficients of extraversion in the predictive interval

Difference! CI: using std. error of the reg. coef.

PI: Using $\sqrt{\text{slope variance}}$, we can calculate predictive interval. → [0.09 – 0.81]

we're interested in PIs!! :)

actually we see that indeed the effect varies quite a lot. It seemed tiny, but then computing PI, we realize that there actually is quite a large variation!!

Contextual models: remember the coffee drinking example

- If the between and within effect differs, we need contextual models to separate them
- In contextual models, we include the variable cluster mean centered / grand mean centerd at the lowest level (the within effect), and the cluster averages at the second level (the between effect)

on top of that, we include the same predictor again in the model, but then as the cluster averages!! to separate within & between effects

  - Explanatory variables X grand mean centered plus cluster means: the level 2 regression coefficient γ_{01} represents the difference between the within and between effect → so this is a convinient model if we wanna test the diff. between within & between effects
  - Explanatory variables X centered within clusters plus cluster means: level 1 and 2 γ_{10} and γ_{01} represent the absolute within and between effect.

If we do this within cluster centering, we are not shifting the dist., we've completely changing it. therefore, interpretation becomes really different: if you score 1 point higher relative to your classmates, then you'll increase XX points on the y-scale → Now it becomes "relative" interpretation!

very impo. to include that when you report the results of this within-cluster centering!!

So we include the same variable twice in our models
Once at the lowest level & second time as a cluster averages at the 2nd level.

ICC in three level models

- $\sigma^2_{v_0}$ variance of the third level errors v_{0k} (extra. w/ three-level)
- $\sigma^2_{u_0}$ variance of the second level errors u_{0ik}
- σ^2_e variance of the residual error term e_{tik}

- The 'percentage' variance at the second and third level:

$$\rho_{class} = \frac{\sigma^2_{u0}}{\sigma^2_{v0} + \sigma^2_{u0} + \sigma^2_e} \qquad \rho_{school} = \frac{\sigma^2_{v0}}{\sigma^2_{v0} + \sigma^2_{u0} + \sigma^2_e}$$

But if we want to have the interpretation of "

- Expected correlation" between two measurements from the same class:

$$\rho_{class} = \frac{\sigma^2_{u_0} + \sigma^2_{v_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e}$$

~ then we need to take account the fact that measures come from the same level 2 clusters, by default, also originates from the same level 3 clusters.

13

So we need to include $\sigma^2_{u_0} + \sigma^2_{v_0}$ these two, if we want to calculate the expected correlation.

13

Questions?

14

14