Multilevel analysis

Contextual effects

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

1

Course outline

Lecture 1
- when/why multilevel analysis
- the multilevel regression model

Lecture 2
- Longitudinal model
- Contextual effects

Lecture 3
- Analyzing dichotomous and ordinal data

2

Monday lecture

- Longitudinal data can be seen as measurements nested in persons
- Multilevel analysis provides a very flexible method to analyze longitudinal data. Here:
  - We use π at the lowest level instead of γ in the notation
  - The baseline model + time should be used as benchmark model for calculating R^2

Emmeke Aarts Multilevel analysis - lecture 2 3

3 2) MLM assumes that we randomly sample individual in the upper level & within individual we randomly sample time points. W/ longitudinal data, this is not the case becuz everyone is usually observed w/ fixed occasions. Therefore there is no variance in time points between individuals. → MLM cannot deal w/ this very well... & this is fixed only when Time predictor is added.

Example Monday lecture: multilevel Analysis of GPA Data

1 • Model for calculating ICC, significance of σ_{u0}^2
  - Time not included: $GPA_{ti} = \beta_{00} + u_{0i} + e_{ti}$

2 • Baseline model for calculating R^2 values
  - Time fixed: $GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + u_{0i} + e_{ti}$

3 • With covariates to predict **within** person variability (in addition to time)
  - Adding job: $GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}Job_{ti} + u_{0i} + e_{ti}$

4 • With covariates to predict **between** (i.e., intercept) variability ~ why average GPA differs over the persons
  - Adding sex and job: $GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i + \beta_{02}\,High\_GPA_i + u_{0i} + e_{ti}$

5 • Allow for random time-effect: we allow every person to have diff. trajectory over time → slope of GPA over time can differ between persons
  - Time random: $GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i + \beta_{02}\,High\_GPA_i + u_{0i} + u_{1i}T_{ti} + e_{ti}$

6 • With cross level interaction to predict slope variability ~ try to explain WHY students show diff. trajectory over time
  - Cross level interaction job and sex : $GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i + \beta_{02}\,High\_GPA_i + \beta_{21}Sex_i\cancel{Job_{ti}}\,Time_{ti} + u_{0i} + u_{1i}T_{ti} + e_{ti}$

Interaction between Sex & Job or between Sex & Time...

Emmeke Aarts Multilevel analysis - lecture 2

4

If you work diff. amount of hours, the influence this has differs between male & female.

Example Monday lecture: multilevel Analysis of GPA Data

Model	M1: random intercept		M2: + time		M3: + job status		M4: +high sch GPA & sex	
Predictor	par est	SE	par est	SE	par est	SE	par est	SE
Intercept	2.87	0.02	2.60	0.02	2.61	0.02	2.53	0.03
Time			0.11	0.00	0.10	0.00	0.10	0.00
Job					-0.17	0.04	-0.17	0.02
High GPA							0.08	0.03
sex							0.15	0.03
Var_occ	0.098	0.004	0.058	0.003	0.055	0.002	0.055	0.002
Var_sub	0.057	0.007	0.063	0.007	0.052	0.006	0.045	0.006
Deviance	913.5		393.7		308.4		282.8	
AIC	919.5		401.7		318.4		296.8	

If we don't include Time,
Var_sub is underestimated,
& Var_occ is over estimated.

Now it's corrected.
But if we don't,
we get neg. R^2 becuz
of this underestimation.

Example Monday lecture: multilevel Analysis of GPA Data

Model	M5: time random		M6: + cross level interac	
Predictor	par est	SE	par est	SE
Intercept	2.55	0.02	2.57	0.03
Time	0.10	0.01	0.09	0.02
Job	-0.13	0.02	-0.13	0.02
High GPA	0.09	0.03	0.09	0.03
sex	0.12	0.03	0.08	0.03
Time*sex			0.03	0.01
Var_occ	0.042	0.002	0.042	0.002
Var_sub	0.038	0.006	0.038	0.006
Var_time	0.004	0.001	0.004	0.001
Covar(sub*time)	-0.002	0.002	-0.002	0.001
Deviance	170.1		163.0	
AIC	188.1		183.0	

Today

1) Centering

2) Contextual effects ~ separating within & between effects is achieved by using diff. centering techniques.

Centering

*Takehome msg: intercept γ_{00} becomes meaningful if we use grand-mean centering.
Main effects of interaction also becomes more interpretable → ex) effect of extraversion for someone w/ avg. teacher experience.

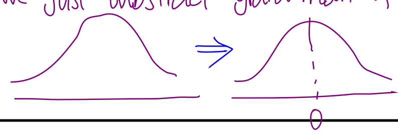
Centering – grand mean

$$popularity_{ij} = \gamma_{00} + \gamma_{10}sex_{ij} + \gamma_{20}extra_{ij} + \gamma_{01}t.exp_j$$
$$+\gamma_{21}extra_{ij} * t.exp_j + u_{0j} + u_{2j}extra_{ij} + e_{ij}$$

- DIY:
  - what does the intercept γ_{00} represent? Predicted popularity when all the predictors are zero; predicted popularity for boys w/ extra = 0 w/ exp = 0...
  - Is it meaningful?
  - How could it be made meaningful?
  - What about the main effect of extraversion and the main effect of teacher experience?

Then it's not really meaningful becuz we know that extraversion ranges from 1 – 10. & Teacher exp. ranges from 2 – 20..sth. So this (0) value doesn't appear in the data so we're centering these continuous variables ⇒ γ_{00} then represents predicted popularity score for boys w/ average extraversion & w/ average teacher exp.

*Grand mean centering: we just substract grand mean of (for ex) extraversion from all the obs. scores



which means, we just shift the entire dist.
So, the center is at diff. location.

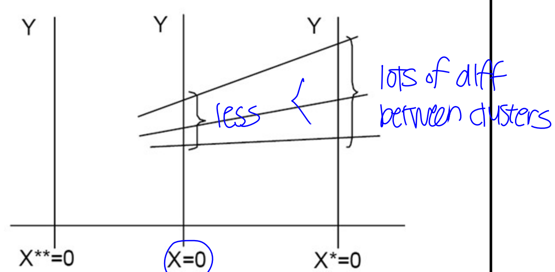
Centering – grand mean

$$GPA_{ti} = \beta_{00} + \beta_{10}T_{ti} + \beta_{20}Job_{ti} + \beta_{30}HighGPA_{ti} + \beta_{20}Job_{ti} + \beta_{01}Sex_i$$
$$+\beta_{21}Sex_iJob_{ti} + u_{0i} + u_{1i}T_{ti} + e_{ti}$$

* centering serves additional purpose if we introduce slope variance into the model:
so if we have both intercept & slope variance, the amount of intercept variance depends on where the intercept actually is.

Note on centering and slope variance

- If the model contains slope variance, the variance of the intercept reflects variance at x = 0
- If we center a variable, we make sure 0 corresponds to a meaningful point, hence the intercept variance has a meaningful interpretation



" the point where X=0 " we shift this by using centering

Emmeke Aarts Multilevel analysis - lecture 2 11

11

* So if we center the data — Not only that γ_{00} has meaningful interpretation, but also, the intercept variance around the fixed intercept has a meaningful interpretation. Becuz it's the amount of difference between clusters that we expect when we have avg. extraversion scores.

Note on centering and (cross-level) interactions

~ centering also has added value on interactions.

Interactions:

- include both "main effects", even if they are not significant
- interpretation "main effects": expected value of the regression slope when the other variable is zero

Now, if we do not center variables, this can become wierd..

Emmeke Aarts Multilevel analysis - lecture 2 12

12

If we include cross-level interaction, all of sudden, the coefficients change drastically... WHY?

NO 0 value in actual dataset — It sort of extrapolates what's going on..

extraversion → expected effect on popularity when teacher exp = 0 &

teacher exp. → expected effect on popularity when extraversion = 0

→ So if we donot center, & uses this 0 point, the main effects also become not interpretable.. they don't carry any meaning...

Example (without centering)

Model:	M$_4$: with random slope			M$_5$: cross-level interaction		
Fixed part	Coefficients	(SE)	*p*-value	Coefficients	(SE)	*p*-value
Intercept	0.74	(0.20)	.001	-1.21	(0.27)	.000
Gender	1.25	(0.03)	.000	1.24	(0.04)	.000
Extraversion	0.45	(0.02)	.000	0.80	(0.04)	.000
Teacher exp.	0.09	(0.01)	.000	0.24	(0.02)	.000
Extra*texp				-0.02	(0.00)	.000
Random part						
σ^2_e	0.55			0.55		
$\sigma^2_{u_0}$	1.30		.000	0.48		.000
$\sigma^2_{u_1}$	0.03		.000	0.00		.042
$\sigma^2_{u_{01}}$	-0.19			-0.03		

Interpretation main effect Extraversion in M$_4$ and M$_5$?

Emmeke Aarts Multilevel analysis - lecture 2 13

13

W/ Centering, the coef. barely changes!

This is why "centering" is a really good idea in ML analysis for multiple reasons:

1) It gives the fixed intercept γ_{00} meaningful interpretation

2) It makes intercept variance meaningful

3) When we include interactions into our model, the main effects also have much more intuitive & interpretable!

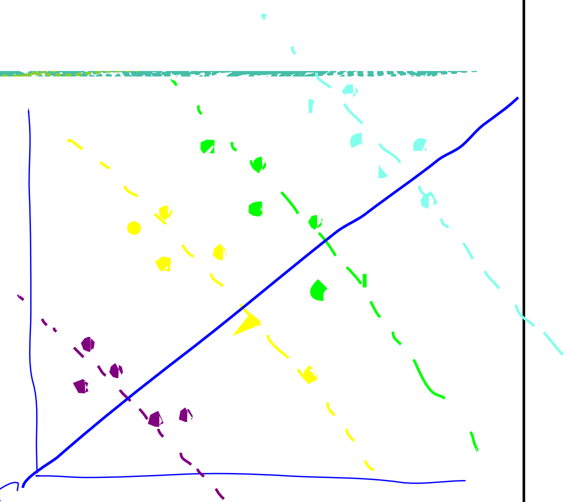
Example (with centering)

Model:	M$_4$: with random slope			M$_5$: cross-level interaction		
Fixed part	Coefficients	(SE)	*p*-value	Coefficients	(SE)	*p*-value
Intercept	4.40	(0.06)	.001	4.37	(0.06)	.000
Gender	1.25	(0.04)	.000	1.24	(0.04)	.000
Extraversion	0.45	(0.02)	.000	0.45	(0.02)	.000
Teacher exp.	0.09	(0.01)	.000	0.10	(0.01)	.000
Extra*texp				-0.02	(0.00)	.000
Random part						
σ^2_e	0.55			0.55		
$\sigma^2_{u_0}$	0.28		.000	0.29		.000
$\sigma^2_{u_1}$	0.03		.000	0.00		.042
$\sigma^2_{u_{01}}$	-0.01			-0.00		

Emmeke Aarts Multilevel analysis - lecture 2 14

14

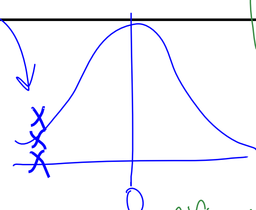
Contextual effects



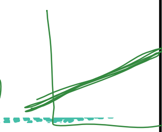
If this is our data then this cannot happen. Becuz that would mean that all the deviations are somewhere here

"the deviation between overall slope γ_{00} & person level slope has to be normally distributed w/ mean of 0 $u_{1j} \sim N(0, \sigma^2)$"

They are forced to have a normal dist. So this is not gonna happen..



person-specific deviations are forced to be in $\sim N(0, sd^2)$ distribution. → So → you will get sth. like this



which is actually a mix of between effect & within effect. As seen here, in some instances the within effect quite differs from Between effect. & If we wanna give an adequate summary of data, it's impo. that we tease those two apart!

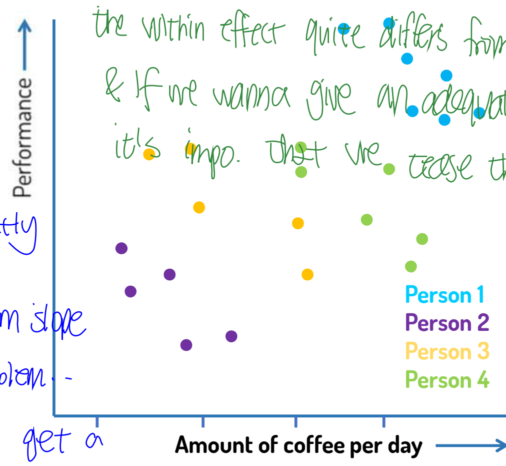
Contextual effects

- DIY: would a multilevel model adequately summarize these data? Why?

Q. Is there a way that MLM can adequately summarize our data?

∴ We need to do an extra trick in order to correctly summarize this data using ML analysis.
Cuz just using ML relying on overall slope & random slope is not gonna do the trick for this type of problem.

What will happen w/ this type of data.. ∵ we will get a sort of mixed like in between the between effects & within effects.



This is where contextual effect comes in!

And this extra trick is = contextual effects

Contextual effects - Frog pond hypothesis

effect of intelligence on school success is not necessarily dependent on absolute value of intelligence, but on your relative intelligence compared to your peers

- Educational science
  - pupils of average intelligence in a class with very smart pupils may find themselves unable to cope, and give up.
  - pupils of average intelligence in a class with very unintelligent pupils, are relatively smart, and may become stimulated to perform really well.
- The frog pond hypothesis states, that the effect of intelligence on school success depends on the relative standing of the pupils in their own class.
  - Apply group mean centering: $intelligence_{ij} - \overline{intelligence}_j$

If you wanna correct for this, you need to apply cluster-mean centering: int. of person i in class j − avg. int. of class j

(this relative measure to his/her class might be more predictive than the absolute value of intelligence.

Looking at relative standing of int. within a class, then actually the effect of int. is quite large

is the "impact of individual characteristics & if varies according to the context in which a person finds him/herself."

So when we do this, we're usually interested in the impact of individual variables on a response within a cluster → "within effect"

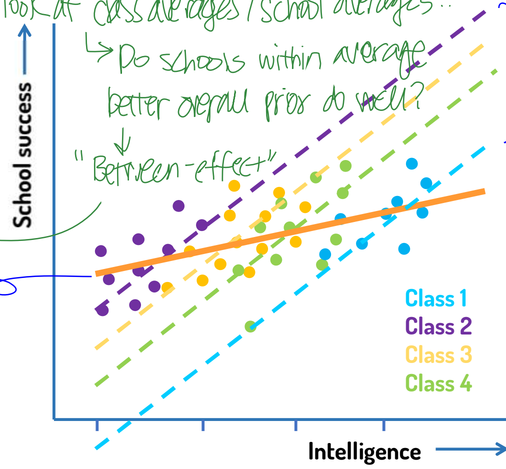
Contextual effects

But also, we're interested in the impact of the contextual variable on the response, so if we just look at class averages / school averages...

↳ Do schools with average better overall prior do well?

"Between-effect"

Here again we see that between-effects, the avg. intelligence ⇒ on school success gives quite a small effect, but the relative standing within a class then actually intelligence affects school success quite a lot!



Within the class, the effects of int. on school success is actually pretty strong. → If we correct for which type of class you're in, then there's actually quite strong relationship between int & school succ.

If we just combine everything, you get the between effect: "how is the avg. int. relates to the avg. school success within a class?" → you see here, quite a low effect. there isn't a very strong relationship between class avg. int & class avg. school success. Overall effect of IT on school success is actually not that pronounced.

We can also look at interaction! "In schools w/ better prior average do children w/ average prior attainment do better /worse?"

*within effect: avg. effect that we see within classes → how is student's relative intelligence relates to school success within a class?

*between effect: how is "avg." class intelligence related to "avg." class success?

You can imagine both can answer diff. research questions.

Contextual effects

- Contextual effect: the impact of an individual's characteristic varies according to the context within which the person lives/works/plays.
- Substantive interests:
  - Impact of the individual variable on response, e.g. good (= above average) prior attainment leads to good performance -> **within effect**
  - Impact of the contextual variable on the response, e.g. schools with better overall prior average do well -> **between effect**
  - Interaction, e.g. In schools with better prior average, children with higher-than-average prior attainment do relatively worse or better..

So there can be also interaction between within & between effects

maybe to have some kind of accelerating effect or the opposite.

Contextual effects

Example: is the effect of extraversion on pupil level identical to effect of extraversion on class level? → So here we'd like to separate $\underbrace{\text{within}}_{\text{"students within a class"}}$ & $\underbrace{\text{between}}_{\text{"class averages"}}$ effects.

conventional MLM w/ fixed slope & random intercept →

- Total model: $Pop_{ij} = \gamma_{00} + \gamma_{10}Extr_{ij} + u_{0j} + e_{ij}$
- Between model: $\overline{Pop_j} = \gamma_{00} + \gamma_{01}\underbrace{\overline{Extr_j}}_{\text{class avg.}} + u_{0j}$
- Within model: $(\underbrace{Pop_{ij}}_{\text{pop of person } i} - \underbrace{\overline{Pop_j}}_{\text{avg. pop. of class } j}) = \gamma_{10}(\underbrace{Extr_{ij} - \overline{Extr_j}}_{\text{relative Extra.}}) + e_{ij}$ ~ within-cluster centering

relative pop.

Between model: just averages in the class level. We'll just aggregate the data, & predict the "avg. class popularity" by "class avg. extraversion"

Within model: just the diff. between Total model − Between model = Within model

$\underbrace{\text{person pop. score adjusted for the cluster avg. pop.}}_{Pop_{ij} - \overline{Pop_j}}$ can be predicted by $\underbrace{\text{extraversion score corrected for the class avg. extraversion}}$

This is called "Within-cluster centering"

$Extr_{ij} - \overline{Extr_j}$

Contextual effects

- Total model: $Pop_{ij} = \gamma_{00} + \gamma_{10} Extr_{ij} + u_{0j} + e_{ij}$ "Mix of between & within effects"
- Between model: $\overline{Pop_j} = \gamma_{00} + \gamma_{01}\overline{Extr_j} + u_{0j}$
- Within model: $(Pop_{ij} - \overline{Pop_j}) = \gamma_{10}(Extr_{ij} - \overline{Extr_j}) + e_{ij}$

- By estimating the Total model we implicitly estimate a Between model and a Within model
- We assume that the effect $\underbrace{\gamma_{10}}_{\text{within effect}}$ and $\underbrace{\gamma_{01}}_{\text{between effect}}$ are equal → we're forcing them to be the same!!
- BUT: this does not need to be the case, γ_{10} and γ_{01} can even have opposite signs!

↳ As we saw in the coffee example, they can be even opposite signs. And it's really impo. that we tease them apart!

⇒ Solution is: we make a Within model where we're predicting students' popularity score corrected for class avg. pop. score by within effect, which is connected to the student level extraversion corrected by class avg. extraversion.

& we have a Between model where we predict the avg. popularity within a class by an intercept γ_{00}, the mean popularity over all the classes and the effect of class avg. extraversion on popularity + a random intercept: u_{0j} → so every class deviation to the unexplained variation in popularity.

Contextual effects

- Can be a serious problem -> coffee example and frog in pond theory
- Solution by Rabe-Hesketh and Skrondal (2005):

- Within model: $(Pop_{ij} - \overline{Pop_j}) = \gamma_w(Extr_{ij} - \overline{Extr_j}) + e_{ij}$
- Between model: $\overline{Pop_j} = \gamma_{00} + \gamma_b\overline{Extr_j} + \underbrace{u_{0j}}_{\text{class-specific deviation}}$
- Total model: $Pop_{ij} = \gamma_{00} + \gamma_w(Extr_{ij} - \overline{Extr_j}) + \gamma_b\overline{Extr_j} + u_{0j} + e_{ij}$

We include predictor (Extr.) twice, once as the relative version & once as the cluster avg. that's fixed over students

Now if we combine these two models to get this contextual effects model, we get the following:

we're predicting $\underbrace{\text{pop. of student } i \text{ in class } j}_{Pop_{ij}}$, by $\underbrace{\text{overall intercept}}_{\gamma_{00}}$ & within effect, which is driven by $\underbrace{\text{corrected version of Extra score: within cluster centered extra}}_{\gamma_w(Extr_{ij} - \overline{Extra_j})\text{: raw extra score} - \text{class avg. extra score}}$

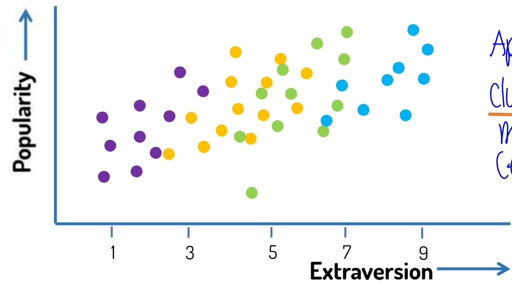
& $\underbrace{\text{between effect, which is related to the cluster mean of extra}}_{\gamma_b \cdot \overline{Extra_j}}$ + $\underbrace{\text{random intercept}}_{u_{0j}}$ + $\underbrace{\text{residual term}}_{e_{ij}}$

Contextual effects

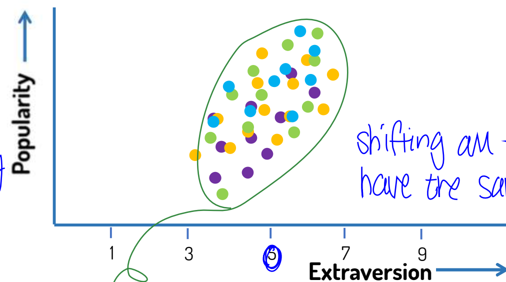
Total model: $Pop_{ij} = \gamma_{00} + \gamma_w(Extr_{ij} - \overline{Extr_j}) + \gamma_b\overline{Extr_j} + u_{0j} + e_{ij}$

- $(Extr_{ij} - \overline{Extr_j})$: individual extraversion score adjusted for cluster mean → so we're using within-cluster mean centering

Cluster mean centering (i.e., within cluster centering)



Apply cluster mean centering →



shifting all the clusters to have the same mean..

So the whole clouds overlap...

All the mean values of each cluster are identical to each other.

& this is all relative standings.

The idea here is: we eliminate the differences between the clusters' averages.

Emmeke Aarts Multilevel analysis - lecture 2 23

23

Contextual effects

Total model: $Pop_{ij} = \gamma_{00} + \gamma_w(Extr_{ij} - \overline{Extr_j}) + \gamma_b\overline{Extr_j} + u_{0j} + e_{ij}$

- $(Extr_{ij} - \overline{Extr_j})$: individual extraversion score adjusted for cluster mean
- γ_b represents the contextual effect : between cluster effects → what's the effect on your outcome if you find yourself in a class w/ higher than average extraversion, for example.
- In the model: same explanatory variable X 2 times in the model.
  - At the individual level (i.e., lowest level): X_{ij} Cluster-mean-centered
    AND
  - At the cluster level: Cluster means X_j

So we have the same variable 2 times in the model { once at the lowest level! once at the upper level!

Emmeke Aarts Multilevel analysis - lecture 2 24

24

between effect = within effect + diff. between within & between effects

Test difference between and within effect

Often, we would like to know if β_w is equal to β_b → If that's the case, we don't need to bother separating this whole thing & doing the contextual analysis. effect

We can reparameterize the model to test his:

- Within model: $(Pop_{ij} - \overline{Pop_j}) = \gamma_w(Extr_{ij} - \overline{Extr_j}) + e_{ij}$
- Between model: $\overline{Pop_j} = \gamma_{00} + (\gamma_w + \gamma_{diff})\overline{Extr_j} + u_{0j}$ +
- Total model: $Pop_{ij} = \gamma_{00} + \gamma_w Extr_{ij} + \gamma_{diff}\overline{Extr_j} + u_{0j} + e_{ij}$

now this γ connected to the cluster mean of extra. represents "Difference" between Within & Between effects!

- Test whether $\gamma_{diff} = 0$ is equivalent to testing whether the within effect and the between effect are equal
- In model: explanatory variables X_{ij} grand mean centered and cluster means X_j

We still have X (Extraversion) 2 times in the model at the upper-level as a cluster means but instead of using cluster-centering, we're just using grand mean centering here! So we are not correcting for cluster-means, we're just grand-mean centering.

So the γ connected to $Extra_j$; cluster mean, represents difference between within & Between effects!

If we operationalize Extra. at the lowest level as a cluster-corrected Extraversion, cluster-relative Extraversion, then each γ represents Within effect & Between effect respectively!

*So what this parameter represents depends on how we put Extra. at the lowest level into the model.

Example

- Example: dataset on Bangladeshi women, the B. fertility study
- Women in communities; we expect women in same community to have similar fertility due to factors like attitudes towards contraception, access to family planning services and SES.
- Variables
  - Community
  - CEB, number of children ever born
  - Age (centered around mean age, which is 28.8)

(Fertility = # of children)

1) Given that you're avg. age, you're expected to have 4 children.
& Every one year older, you get about a quarter (1/4) child extra.
& w/ community of older age, this effect is expected to be less → for ex) you're in a community of 1 year older in avg. age, then you're expected to have 1/5 ...

Example

- If we assume within and between model are equal: then we only estimate one reg. parameter for age.

$$Fertility_{ij} = 3.94 + 0.24\ age_{ij} + u_{0j} + e_{ij}$$

- Age (grand mean centered) and Mean-age : If we reparameterize the model to separate the Within & Between effects, this is the

$$Fertility_{ij} = 3.94 + 0.24\ age_{ij} - 0.05\ \overline{age_j} + u_{0j} + e_{ij}$$

is not corrected for age, we find that the person model. within the cluster belongs to.

1) DIY: what does the above model tell us?

- Cluster mean centered Age and Mean-age:

this now becomes the relative age.

$$Fertility_{ij} = 3.94 + 0.24\ age_{ij} + \cdots \overline{age_j} + u_{0j} + e_{ij}$$

2) DIY: what should be the value of the regression coefficient for $\overline{age_j}$?

∴ 0.24 + (−0.05) = 0.19 ∼ If we operationalize in one way, then the two γs represent the absolute within effect & absolute between effect.
BUT, if we parametrize in another way, then the first γ represent the absolute within effect & the other γ represents the Diff between within & Between effects.

This model can be used to test whether there's actually difference between Within & Between effects.

Emmeke Aarts Multilevel analysis - lecture 2

27

Summary

- Grand mean centering aids interpretation of (cross-level) interactions, the intercept and intercept variance when slope variance is present
- If the between and within effect differs, we need contextual models to separate them
- In contextual models, we include the variable cluster mean centered at the lowest level (the within effect), and the cluster averages at the second level (the between effect)
- When we want to test whether the within and between effect significantly differs, we reparametrize the model such that the between effect represents the difference between the within and between effect, by at the lowest level entering the variable grand mean centered instead of cluster mean centered

Emmeke Aarts Multilevel analysis - lecture 2 28

28