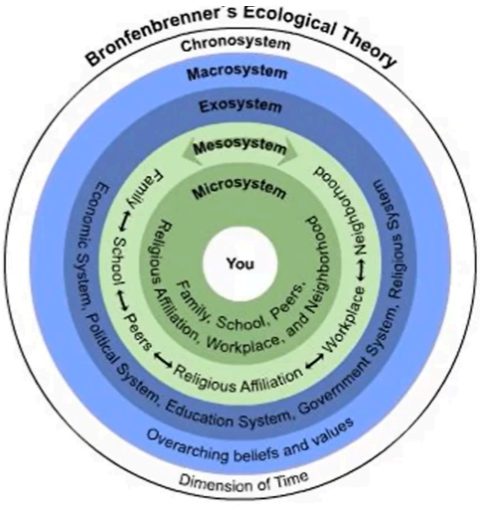
Week#1:
Mixed-effect Models

One category is Multi-level models:
In multilevel models, your observations are nested in each other.



exp: systems nested in each other
This is quite similar to where Ps were nested within larger units (such as schools, classrooms, etc.)
***nested/hierarchical/multi-level structures

Örn: 2 level hypothetical model; we are investigating involvement in crime. We will sample adolescence from diff. neighborhoods.

Micro level
The association between impulse control and crime involvement in adolescence

Macro level
The association between neighborhood SES and crime rate

Macro-micro interactions
Increased association between impulse control and crime involvement for low-SES neighborhoods

Inds level (level 1, associations with crime at the Ps level) and neighborhood level (level 2, associations with crime at neighborhood level)

**so here we can look at whether the neighborhood characteristics predict our level 1 outcome (the adolescents' crime involvement)

***our outcome measure should be at the level 1, so our level 1 should include the outcome measure. But, higher levels can include predictors.
E.g. our predictor at level 1: impulse control; predictor at level 2: neighborhood SES

***Macro-micro interactions: can it be the case that impulse control predicts crime involvement for low SES neighborhoods (so in high SES neigh. your level of impulse control will not much affect your crime involvement, but in low SES neighb. Your level of impulse control will predict your crime involvement)

Why do we need multilevel modeling?: it allows for testing our theories more precisely (theoretical specification); it also helps with correctly specify random factors (this reduces type1 error)
Why does it help to reduce type 1 error?--> all the other models, tests we used so far assume that all the Ps are independent from each other (because the error terms of diff. Ps should be meaningfully related to each other).
Multilevel model (when we have more than 1 random factor), you violated independence. (because when you have multiple levels, multiple random factors, the error terms of diff. Ps from the same level might be more correlated with each other)
e.g. adolescents' error terms from neighborhood A may be more correlated with each other than neighborhood B. (because those from the neighborhood A share the same environment etc. their error terms are more closely correlated with each other)
So, LEVELS CREATE DEPENDENCY IN THE DATA (so your observations are not independent anymore)
Higher that dependence, the more important recognizing the micro level becomes. (the higher the type 1 error rate if you ignore it)
Higher dependence means that; the macro level explains a more important portion of the variability in your outcome variable. (so ignoring that dependency can highly increase type 1 error)
Thats why we should be very careful with identifying random factors in multilevel models.

Generalization to the population of groups: now when you have more than 1 random factor, you have more than 1 group to generalize your findings. (in the exp: generalizability to the indv. And also generalizability to neighb.) So here we are testing both ind.level effects (impulse control-crime rates associations) and group-level effects (neigb SES- crime rate associations).
So are my findings generalized to all levels? That again depends on correctly **identifying the random factors**.

How can we estimate the degree of dependence:
(E.g. how can we estimate the role of neighborhood in explaining variability in crime rates)
We will compute **Intraclass Correlation (ICC)** for this → a measure of the degree of dependence or similarity across individuals (a measure of homogeneity)
(exp: it can be the degree of dependence coming from; monozygotic twins, repeated measure Of the same person, residents of the same neighb.)

The ICC is proportion of total variance due to differences between groups

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

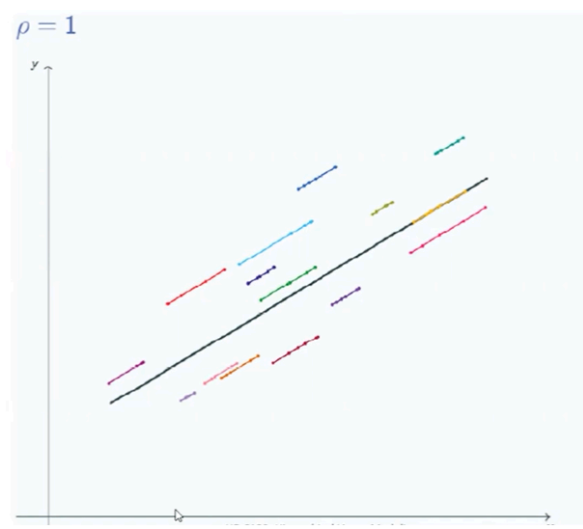
The ICC = 0 if no group effects and 1 if no within-group differences. (u= level 2 error, e= level 1 error)

Here we will divide variability into 2 sources: 1) variability due to inds. 2)variability due to group levels → ICC is variability due to groups /total variance of these 2 sources

If ICC is 0=then you dont have a 2nd random factor
If ICC is 1=there is no within-group difference.(means that everything is coming from 2nd factor and no effect coming from ind level)
***these are hypothetical; in reality almost always never ICC= 1 or 0, it is in btw 0-1)

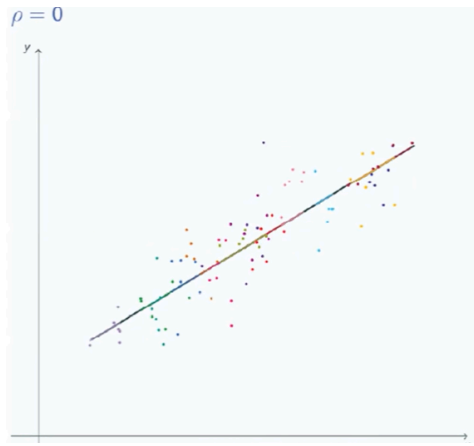
X-axis: impulse control problems, y-axis: crime involvement, colorful lines represent diff neighb., long black line represents overall regression line (average for all Ps)

If ICC =1: means that if you know the nighb. of the the person you can model their association btw impulse control and crime involvement (everything is explained by the negihb.)
Here also as represented by the image; the data is quite clustered; so everybody behaves as their neigh. would predict.

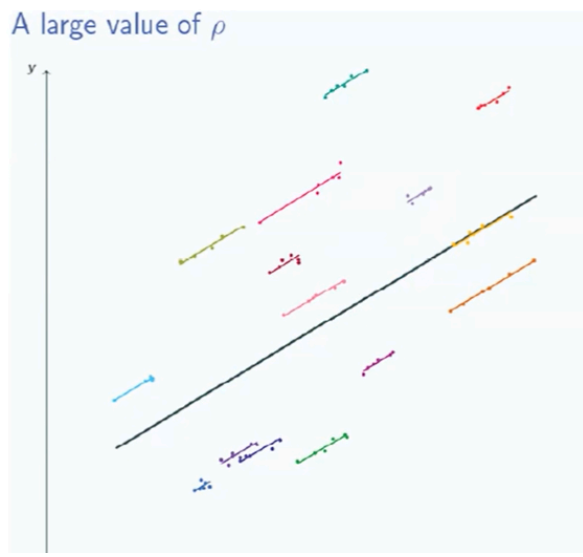


X-axis: impulse control problems, y-axis: crime involvement, colorful lines represent diff neighb., long black line represents overall regression line (average for all Ps)

If ICC=0; means that a single regression equation would work just fine (because neigh. Does not explain any variability)

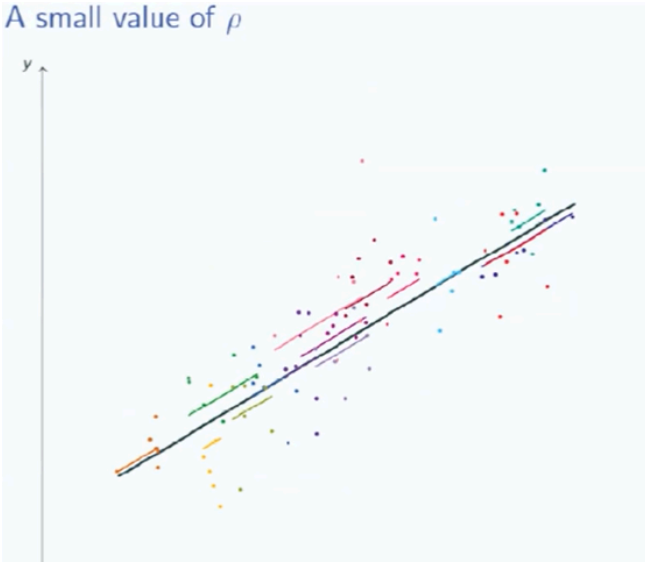


But these are hypothetical cases. In reality our model would look like this:



A large value of ICC:

The data is still clustered, so the differences are btw-groups (Ps in the same neighb. are more correlated with each other than with those from other neigh.; within-neigh differences are smaller) but there are still ind. Variations.



A small value of ICC:
Means that; btw-group differences (differences btw neigh.s) are not that prominent, rather within-group differences are much prominent (the differences btw individuals in the same neigh)

But even in very small values of ICC, you should stick with multi-level models

Sum table of increase in type 1 error rate:

	ρ		
N_i	.01	.05	.20
10	.06	.11	.28
25	.08	.19	.46
50	.11	.30	.59
100	.17	.43	.70

- Ni: num of groups (neighb.) in our data

If ICC (dependency based on group membership) is .20 ; %20 of variability is due to your group membership (%20 of crime involvment is due to which neighb. you live)

No matter how many neigh. you sample, if you ignore neighb. En sağdaki colum is type 1 error rates. So if you sample 100 neigh. And ICC is .20 ; your actual type 1 error rate is .70 (which is very high)
But even when ICC is small (.01) and num of neigh you sample is small (10); even in that scenario alpha is .06

So, increase in type 1 error occurs pretty quickly when you have many random factors!!! So identifying and properly modeling random factors are imp.

Data exp:

	bar	smile	beer
1	a	8.647	0.524
2	a	11.296	2.174
3	a	12.292	2.491
4	b	12.633	2.015
5	b	11.197	1.689
6	b	9.203	1.026
7	b	11.270	3.261
8	b	9.376	2.426
9	b	10.091	1.477
10	b	9.416	2.100
11	b	9.566	0.333
12	b	12.346	2.546
13	b	10.344	2.866
14	b	11.878	2.250
15	b	9.298	0.519
16	b	11.118	2.792
17	b	12.734	3.167
18	c	8.323	0.527

How to decompose total variability into 2 sources: 1) variability due to inds. 2)variability due to group levels????

Data: investigating the association between drinking behaviors and smiles produced. They picked 15 diff bars and measure how much beer ppl drink and how much they smile (outcome var.)

This data has a multi level (nested/clustred) structure : you have Ps nested within diff. bars
So we use a linear mixed model (we will separately estimate variability due to Ps and due to bars)

!!!In a null model (no predictors involved yet): we have 3 parameters (level1 , level2 parameters and intercept (referring to the grand mean of the outcome variable))

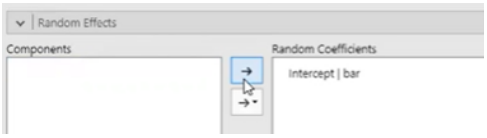
***Decomposing our outcome (smiles) variability into 2 sources: bars and inds.

We can compute the average within each bar and then get the mean of those averages (the grand mean). =this is what the mixed effects models will do → first the software will run the analysis for each cluster and then get an average of those analyses.
(basically, it will allow each bar to have its own mean) → for this we need to allow for random intercepts (intercept refer to the mean level of the outcome variable assuming that all predictors are 0 and if there are no predictors intercept will be the grand mean)

In mixed models: instead of imposing a single intercept to each bar, it will allow each bar to have its own mean → thats why this model is called **random intercept model**

Jamovi → Linear Models → Mixed Model

Define Dv, Factors, Cluster variables (in this case it is bars, Ps are nested within bars)
Define fixed and random effects



(defining a random intercept term: allow for each bar to have its own separate intercept)

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Name	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	7.875	0.467	6.960	8.790	13.829	16.871	< .001

Random Components

Groups	Name	SD	Variance	ICC
bar	(Intercept)	1.763	3.109	0.644
Residual		1.310	1.717	

Note: Number of Obs: 234 , groups: bar 15

Look at the fixed effects parameter table; the intercept term refers to the grand average in smile behav.
Look at random components table: shows random variability in smile behav due to bars and due to inds. (here we are decomposing the total variability into its sources)
ICC: is the proportion of variability due to level 2/total variability (ICC: .64 =3.10 /(3.10+1.71))
This ICC tells us that; taking into account the clustering factor (level 2) is very critical because a sizable portion of the variability is coming from level 2 variable (bars).

Week#2: Level 2 predictor & Model Equations:

A special sub-category of Mixed effect models: Multilevel models (distinguishing feature of these models is that one source of variation is nested within the other source. The idea is to see whether these two levels explain diff amounts of variability in your outcome variable)

Exp:

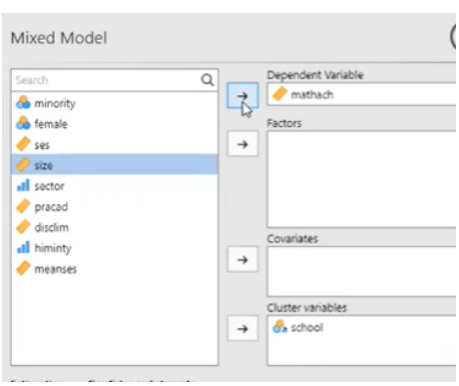
	school	minority	female	ses	mathach	size
1	1224	other	1	-1.528	5.876	
2	1224	other	1	-0.588	19.708	
3	1224	other	0	-0.528	20.349	
4	1224	other	0	-0.668	8.781	
5	1224	other	0	-0.158	17.898	
6	1224	other	0	0.022	4.583	
7	1224	other	1	-0.618	-2.832	
8	1224	other	0	-0.998	0.523	
9	1224	other	1	-0.888	1.527	
10	1224	other	0	-0.458	21.521	
11	1224	other	1	-1.448	9.475	
12	1224	other	1	-0.658	16.057	
13	1224	other	0	-0.468	21.178	
14	1224	other	1	-0.988	20.178	
15	1224	other	0	0.332	20.349	
16	1224	other	1	-0.678	20.508	
17	1224	other	0	-0.298	19.338	

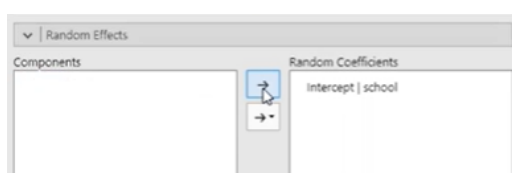
Students nested within schools. We have student level and school level. Outcome variable is math achievement. (our outcome variable will be always at level 1)
We have student level predictors: student SES, gender, minority statements etc.
And school level predictors: size of school, school sector, %of students in academic track, etc
So for all levels we have predictors for each level.

***Compute ICC (to decompose total variability into its sources to explain how these levels explain the variability in outcome variable) exp: how much variability in match achievement is due to students and due to schools —> this computation is called a null model.(so no predictors of these levels are involved, we are just looking at how much variability in outcome measure is explained by the diff. levels/ so it decomposes the variability into its sources)
In this null model (also called random intercept model): we allow diff clusters to have its own means and then compute the grand mean of all groups. So the model has only 1 intercept (grand intercept). We allow intercepts to vary randomly across cluster units.

In our exp: the random intercept model will allow our mixed effect model to compute mean math achivem. within each school separately, and then compute a grand mean.

Compute ICC:





Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	12.637	0.244	12.158	13.116	156.647	51.707	<.001

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	2.935	8.614	0.180
Residual		6.257	39.148	

Note. Number of Obs: 7185 , groups: school 160

Grand mean of math achieve: 12.63 (we allowed each of the schools to have their own means; meaning that mean math achievement was estimated → estimating diff means of each schools based on the students math achivm. scores within each school)
So, for X num of schools in our dataset ; we have X num of diff mean math achivem. Estimations.

Variance in math ach. due to schools: 8.61
Variance in math ach. due to students: 39.14

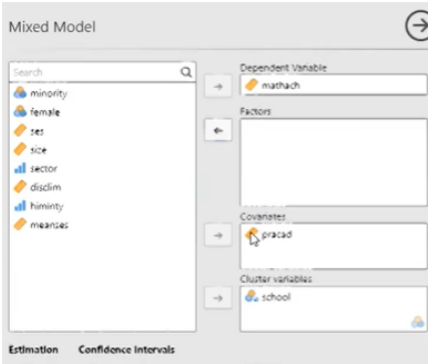
8.61/ (8.61 + 39.14) = 0.18 = ICC

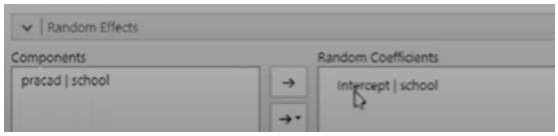
!! If I include school level variables in the model (size of school, school sector, %of students, etc), then variability due to schools will decrease; or If I enter student level variables, variability due to students will decrease.

(REMINDER in the regression framework: each added predictor in the model → the variability I can explain increases and error term (variability I cannot explain) decreases.)

Question: Now that I can explain variability due to diff. levels; whether I can use my predictors to explain some of the variability ? (the more I explain, more reductions in variance)

Exp: lets enter a school level predictor in to the model (proportion of students in the academic track)





NOTE: Where to put the predictor?: if it is a continuous var. Put it in “covariate” box, if it is a categorical var. put it in “factors” box.

NOTE: We estimate random coefficients only for level 1 predictor (whether the level 1 terms of the model randomly vary across level 2 units. academic track var. is already a school level variable, so no need to put it in to the random coefficients box

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
pracad	134.933	1	158.867	< .001

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	12.784	0.180	12.431	13.138	156.026	70.879	< .001
pracad	8.232	0.709	6.843	9.621	158.867	11.616	< .001

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	2.061	4.250	0.098
Residual		6.257	39.144	

Note. Number of Obs: 7185 , groups: school 160

Here we have a predictor (this is not a null model no more)

The more students in academic track in a school, the higher the math achivm. Scores. (positive estimate of 8.23)

After accounting for proportion of students in academic track in the model, the unexplained variability is down to 4.25 (it was 8.61 in our null model)
Thats why our ICC declined. (Here the ICC is .09. (it was .18 without the predictor included)

So by adding the level 2 predictor, I was able to explain more proportion of variance due to schools

Writing 2-level Model Equations:

Intercept-only model

Level 1 (student level)

$$(\text{Mathach})_{ij} = \beta_{0j} + r_{ij}$$

Level 2 (school level)

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

Mixed (combined) model

$$(\text{Mathach})_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Note: an intercept-only model estimates only 3 parameters: grand mean (a fixed effect), student level variability (a random effect) and school level variability (a random effect)

Adding a Level 2 predictor

Level 1 (student level)

$$(\text{Mathach})_{ij} = \beta_{0j} + r_{ij}$$

Level 2 (school level)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Pracad}) + u_{0j}$$

Mixed (combined) model

$$(\text{Mathach})_{ij} = \gamma_{00} + \gamma_{01}(\text{Pracad}) + u_{0j} + r_{ij}$$

Note: when we add a predictor in the model; we now have 4 parameters: intercept (fixed effect), the predictor (fixed effect), student level variability (a random effect) and school level variability (a random effect)

!!!! When you add a level 1 predictor, the num of total parameters may increase more than one.

Week#3: Random Slope Models

Fixed or random slope model??

Adding level 1 predictor in the model (fixed slope):

(How SES (level 1 predictor) is linked with student math achievement)

Adding Level 1 predictors

Level 1 (student level)

$$(\text{Mathach})_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}) + r_{ij}$$

Level 2 (school level)

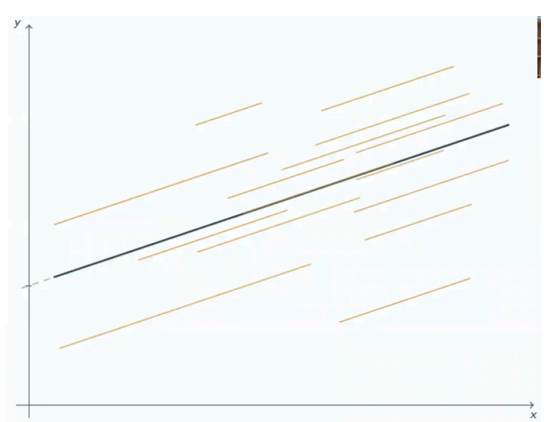
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

Mixed (combined) model

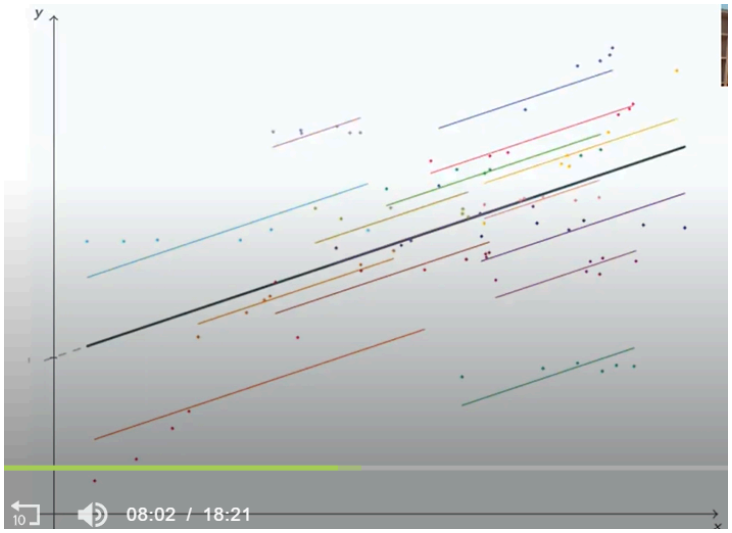
$$(\text{Mathach})_{ij} = \gamma_{00} + \gamma_{10}(\text{SES}) + u_{0j} + r_{ij}$$

If I treat the slope (B1j) here as fixed, then it means I will estimate the same slope for each school. So we add no random component in the equation.

In which cases we use a fixed effect?:



X : SES, Y: math achivm., bold line; fixed slope (the average slope/gama 10), intercept: gama 00
Orange lines; each school having different intercepts. But I wanted the slope to be a fixed effect (the same slope for each school), so thats why each of these orange lines are parallel to the bold line (fixed slope)



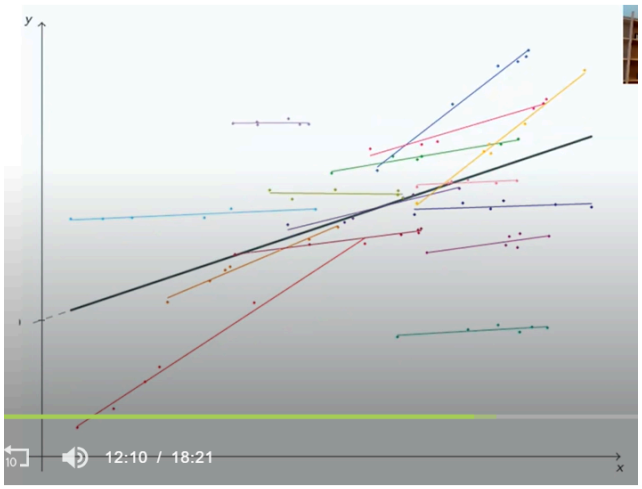
This graph; there is not much heterogeneity in the association btw the predictor and outcome across clustering units.(the averages differ but the slope is similar)
These dots are ind students.
Here, in each school there is a pos. Association of a similar size btw SES and mathachiv.
There is not much heterogeneity, so we do not need to add a random component for the slope term as it wont give much information. So we can keep the slope as a fixed effect

!!! estimating random slope is more costly than estimating a random intercept. Because there are additional parameters you need to estimate.

**models become more complex, as we introduce more parameters and as a result, the available data points may not be sufficient to successfully achieve a solution.
If we are limited by a fewer data points, these available data points may not be enough to estimate all random slopes, In those cases we use fixed slopes.

Sometimes, especially if you have a few level 2 predictors, researchers look at whether they gain anything by adding a random slope term. If you do not gain anything, then you can just drop the random slope term. So keep the slope as a fixed effect, without any random slope term.

But what if we have this example:



Here, the best fitting line for each school is very different than others, for some schools we see a pos. Association, for others we see no association (a flat line)

If this is the case, then you need to add a random slope term in order to reliably estimate the average effect (the average association btw SES and math achivm across schools)

Then our equation must be (a random slope):

Random slope model

Level 1 (student level)

$$(\text{Mathach})_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}) + r_{ij}$$

Level 2 (school level)

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Mixed (combined) model

$$(\text{Mathach})_{ij} = \gamma_{00} + \gamma_{10}(\text{SES}) + u_{0j} + u_{1j}(\text{SES}) + r_{ij}$$

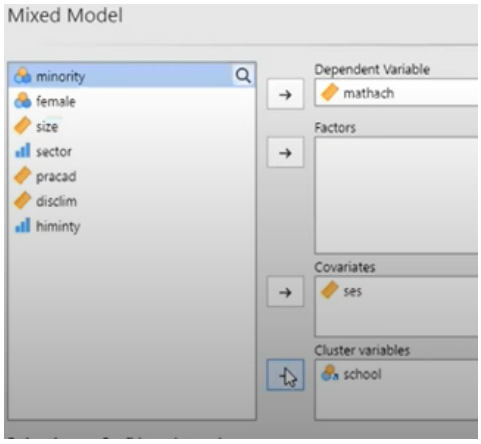
Added u1j (we allow each school to have their own slopes and then we average across those individual slopes to achieve the average slope across schools (gama 10))

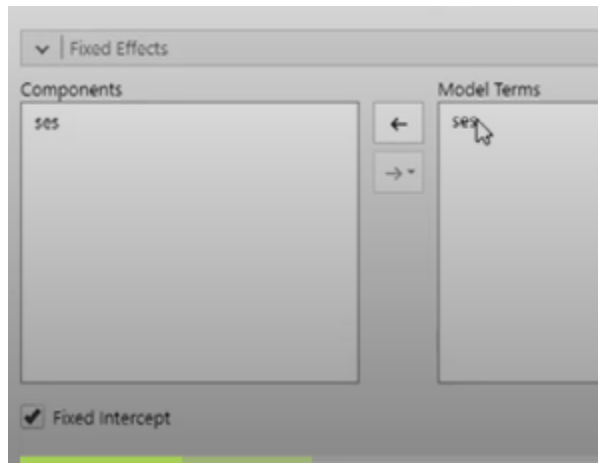
We know have more than one random component (a random term for intercept and for the slope). Thus, we also have a correlation btw intercept and slope. Once we have more than one random component at level 2 we need to estimate the correlation btw them.

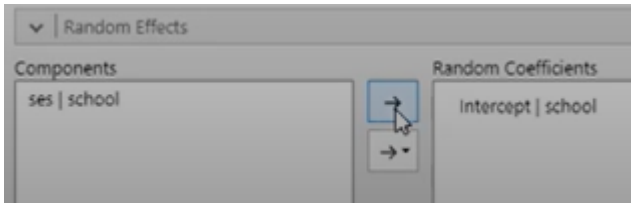
How many parameters do I have in this random slope model?:
→2 fixed effects (intercept and slope), 2 random effects at level 2 (deviation of each school from the grand intercept (u0j) and deviation of each school from the grand slope (u1j)), 1 residual term (a random effect) and 1 correlation (btw the random components).

!!not adding a random error term might increase type 1 error because it may increase SE if stand. random intercept is smaller, SE bigger ((t= gama/SE), type 1 error risk increas(t= gama/SE)without a random slope, there might be a type 1 error increase risk.

Jamovi exp:
*Mixed model predicting math achivm from student SES







(a random intercept model. We dont introduce the random slope of ses yet)--> so for now, we allow schools to have their own mean intercept (mean mathachivm) but we will estimate a single slope value for every school (same association btw SES and math achivm for every school)

NOTE: in mixed models; we have multiple ways of centering:
-one approach is called "grand mean centering" : you subtract the overall mean/the grand mean from each observation (this the jamovi by default uses)
- other method is called "group/cluster mean centering": you subtract the particular cluster's mean score from each observations. Can be very helpful when we model development over time

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
ses	511.164	1	6838.078	< .001

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	12.658	0.188	12.289	13.026	148.302	67.334	< .001
ses	2.390	0.106	2.183	2.597	6838.078	22.609	< .001

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	2.184	4.768	0.114
Residual		6.086	37.034	

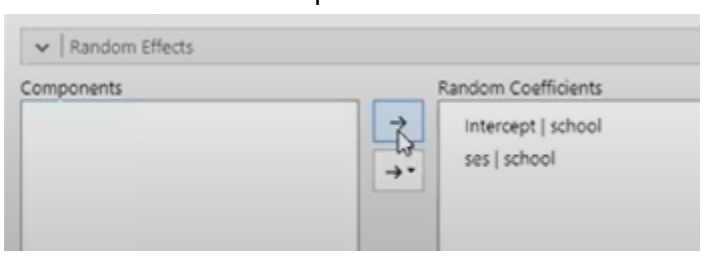
Note. Number of Obs: 7185 , groups: school 160

,

Here, slope (the effect of SES on math achivm) is 2.39 (for 1 unite increase in SES, leads to 2.39 point increase in math achievem)
Intercept; mean math achivem when SES is 0.

Vairbility due to schools: 4.76
Variability due to students: 37.03

İf we add a random component for the school term:



Random slope term → Now we also have variability in the effect of SES on math achivm. So for each school we estimated a different slope and the variance of those slopes is 0.41 (variance of association btw SES and math achivem across schools)

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
ses	410.701	1	157.530	< .001

Note: Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	12.665	0.190	12.293	13.037	145.548	66.715	< .001
ses	2.394	0.118	2.162	2.625	157.530	20.266	< .001

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	2.197	4.829	0.116
	ses	0.643	0.413	
Residual		6.069	36.830	

Since we have more than one random component at level 2, we know have a estimated correlation btw those random components (slope and intercept) (which is -0.1):

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
school	(Intercept)	ses	−0.109

This means; a very low neg. association btw the random intercept and random slope → for schools who have high mean math achivm scores, the association btw SES and math achivm. Is weaker.

Week#4: Random Slope Models cont.

Adding level 1 (student SES) and 2 predictor (school sector) simultaneously:

**if you are interested in the main effects of a level 2 predictor adding it to the intercept equation will be enough. (putting the level 2 predictor in the intercept equation)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Sector}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

**but you may also be interested in whether the level 2 variable effects the slope (whether the association btw student SES and math achievem depends on the school sector); then you should put the level 2 predictor in the slope equation.
***So essentially; adding a level 2 predictor into the slope equation → tests an interaction effect for you.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Sector}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Sector}) + u_{1j}$$

Level 1 (student level)

$$(\text{Mathach})_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}) + r_{ij}$$

Level 2 (school level)

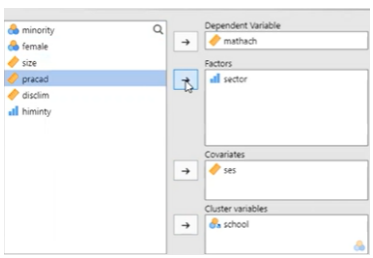
$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Sector}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Sector}) + u_{1j}$$

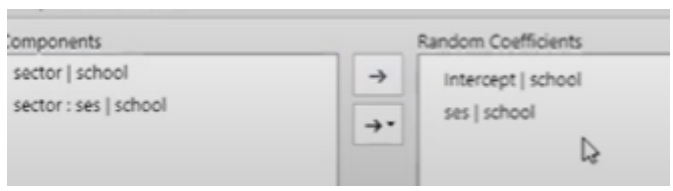
Mixed (combined) model

$$(\text{Mathach})_{ij} = \gamma_{00} + \gamma_{10}(\text{SES}) + \gamma_{01}(\text{Sector}) + + \gamma_{11}(\text{SES*Sector}) + u_{0j} + u_{1j}(\text{SES}) + r_{ij}$$

***Gama 11 (SES *Sector): actually an interaction term; this is called cross level interaction (a higher level variable moderating the effect of lower level variable on the outcome)
(whether the association btw SES and math achievement is moderated by the school sector)

JAMOVI exp:





Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval Lower	Upper	df	t
(Intercept)	(Intercept)	11.472	0.231	11.020	11.927	153.797	49.569
ses	ses	2.385	0.118	2.154	2.616	157.852	20.238
sector1	catholic high school - public high school	2.541	0.345	1.865	3.216	151.306	7.375

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	1.991	3.965	0.097
	ses	0.659	0.434	
Residual		6.066	36.801	

Note. Number of Obs: 7185 , groups: school 160

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
school	(Intercept)	ses	0.550

The intercept is the mean math achievement when the other predictors are 0. In this case 0 has meaningful values. 0 = public highschools for school sector var. and 0=average SES for student SES var.
So, intercept is mean math achivm for average SES students who attend to public highschools
Effect of SES; a positive association btw SES and math achivm. (this reflects the effect of SES on math achievem in public schools (when sector is 0))

Effect of sector; average math achivm higher in catholic schools (2.54 higher) compared to public schools (this reflects the effect of sector on math achivem for students with average SES(when SES is 0))

Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	1.991	3.965	0.097
	ses	0.659	0.434	
Residual		6.066	36.801	

Note: Number of Obs: 7185 , groups: school 160

Looking at random effects:
*In this model we allowed schools to have their own math achieve, and also allowed schools to have their own slope representing the association btw SES and math achievm.

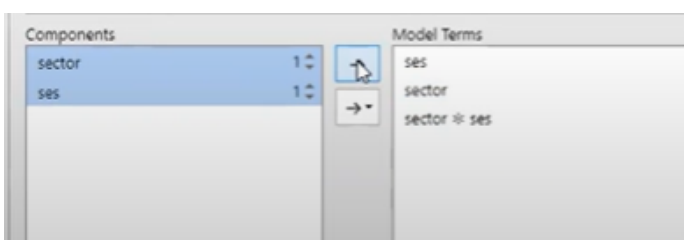
**each time we add a new predictor to the intercept equation, we will explain some variability in the intercept. → the school variance score will go down
**so far in this result table; sector is included only in the intercept equation at level 2, not in the slope equation at level 2 yet. So interaction effect is not yet here.

**when add the interaction effect (when add the sector predictor to explain slope at level 2) → the SES variance score will explain some variability in the slope

**when add the interaction term, if the variance explained score goes down significantly, then we can remove the random effect of SES slope. Because if there is not much variability left to explain, then we can remove this effect.

**deciding on random factors: if your theory suggests that there should be a random factor keep it, but if not be mindful about whether there is variability. If there is not much variability do not keep the effect in the model, just remove it.

-we will decide whether to keep or remove the interaction effect after adding the interaction term:



-the interaction effect is essentially, adding the sector predictor to the slope equation at level 2 (so we are trying to explain some variability in the SES slope; looking at whether the association btw SES and math achivm depends on sector)

Look at the random components again:

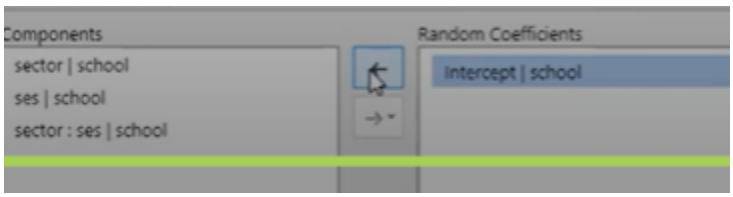
Random Components

Groups	Name	SD	Variance	ICC
school	(Intercept)	1.955	3.822	0.094
	ses	0.275	0.076	
Residual		6.065	36.78	

Note: Number of Obs: 7185 , groups: school 160

The SES variance score declined. So we explained very big amount of variability in the previous table but now it declines critically. Meaning that, by adding sector as a predictor in the model, a sig amount of variability disappears, this means that I do not need to keep this term in my model.

*so remove the SES as a random component:



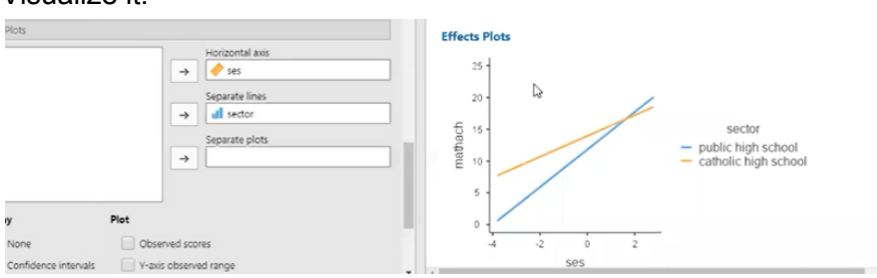
Interpret the findings:

Fixed Effects Parameter Estimates

Name	Effect	Estimate	SE	95% Confidence Interval Lower	95% Confidence Interval Upper	df
(Intercept)	(Intercept)	11.752	0.232	11.298	12.207	152.414
ses	ses	2.958	0.143	2.677	3.238	1358.063
sector1	catholic high school - public high school	2.129	0.346	1.451	2.807	145.356
ses * sector1	ses * catholic high school - public high school	-1.313	0.216	-1.736	-0.891	1397.975

Random Components

- A sig. Interaction term, a negative interaction effect. (interaction term assess whether the SES effect depend on the sector)
- So, SES effect on math achiv sig.ly changes for diff sectors
- For catholic schools, SES decreases math achievm -1.31 (just plug in the numbers; catholic=1, public = 0. If plug the numbers for public school (0) in the equation it will be 0. For catholic, if you plug in 1 in the equation; 1*(-1.31) =-1.31 → it means that for catholic schools the effect of SES on math achivem is decreased by -1.31)
- Visualize it:



***we can also interpret that; sector makes a sig. Difference on math achievem at low SES levels. For high SES level students the sector does not mean any difference on math achivem.

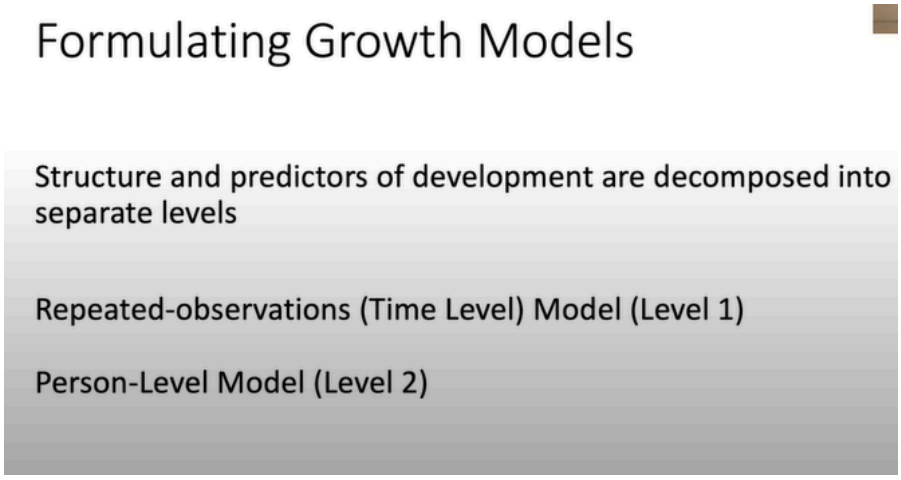
group mean: x-x(groupmean) (0 is average math achieve at school j) 1 unit increase corresponds to ; the school j's average SES change (most of the time we use this; why? each school have diff characteristics that result math achivem scores to differ; take all scores and subtract each schools' SES to that average → we are removing the btw school differences)

grand mean x-u (0 is average math ach overall) ; 1 unit ses increase respect to entire sample's average SES

Growth Curve Models

Analyzing change over time

Within Anova: have constrations (sphericity assumption, no missing values) → not easy to establish in longitudinal data



Diff types of growth models: linear growth model, quadratic growth model, cubic growth model

Linear Growth Model

Has one Level-1 predictor, TIME

Has two Level-1 parameters: (i) intercept (status measure) and (ii) slope coefficient (growth rate)

Interested in Time effects on outcome variable

Intercept (mean level of outcome var. when other predictors are 0; status of that behv at time 0 (time 0 can be anything you want, center it smt that makes the most theoretical sense)) and slope of time (your growth rate; change in outcome behav as one unit increase in time)

Brief note on notation…

Level	Coefficient	Subscript	Error
Repeated-measures	π	t	e
Individual	β	i	r
Organization	γ	j	u

Specifying the Linear Growth Model

Step 1: The fully unconditional model

Level 1

$$Y_{ti} = \pi_{0i} + e_{ti}$$

What does the fully unconditional model tell you?

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

Full unconditional mode: gives ICC value (the amount of variability at the btw person level)
(1-ICC): tells us variability at the within person lvl
You can decompore btw-person and within-person variability → how much change in behav due ind being diff than each other and how much due to person 1 developing over time

Within-person designs: can look at both within & btw effects

If not much within-person variability then the model may not be helpful because the behav dosent change over time anyway. Really the diff are at btw-person

Specifying the Linear Growth Model

Step 2: The unconditional growth model

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i}(Time_{ti}) + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$

What does the unconditional growth model tell you?

Pi1i: critical information about the amount of change in behav as a result of 1 unit change in Time

Why add random slope term (r1i): many psycal behav development shows ind diff. So we allow the effect of time to vary across inds

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i}(Time_{ti}) + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

The interpretation of the intercept depends on how you center *Time*

Make sure your time variable has a meaningful 0 point

Note: Usually the manipulations are at the btw-person level

Step 3: Add the remaining substantive predictors

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i}(Time_{ti}) + \pi_{2i}(X_{ti}) + e_{ti}$$

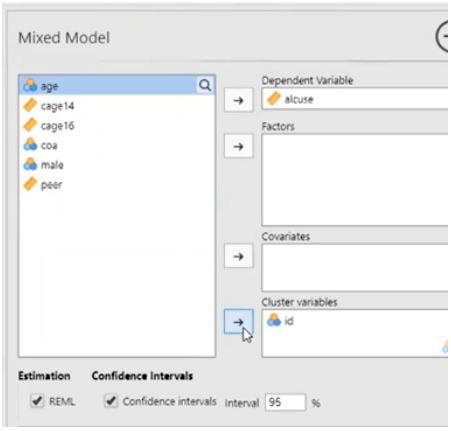
Level 2

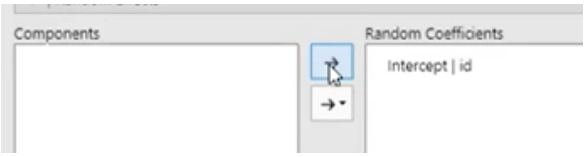
$$\pi_{0i} = \beta_{00} + \beta_{01}(Z_i) + r_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11}(Z_i) + r_{1i}$$
$$\pi_{2i} = \beta_{20} + r_{2i}$$

Measure additional predictors (Xti) at level 1

At level 2: Zi →model ind diff factors (how do ind diff effect development/how can we explain differences in development), tell me whether these ind diff/person-level predictors effect the rate of development (Pi1)

Jamovi exp:
Data: alcohol use; adolescence followed 3 times (per year, 3 years)
*in this dataset, time (time variable here is “age”) is centered at the starting point of the study (when Ps are at age 14)
*So the intercept will tell me the mean amount of alcohol use at age 14





Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval		df
			Lower	Upper	
(Intercept)	0.922	0.096	0.733	1.111	81.000

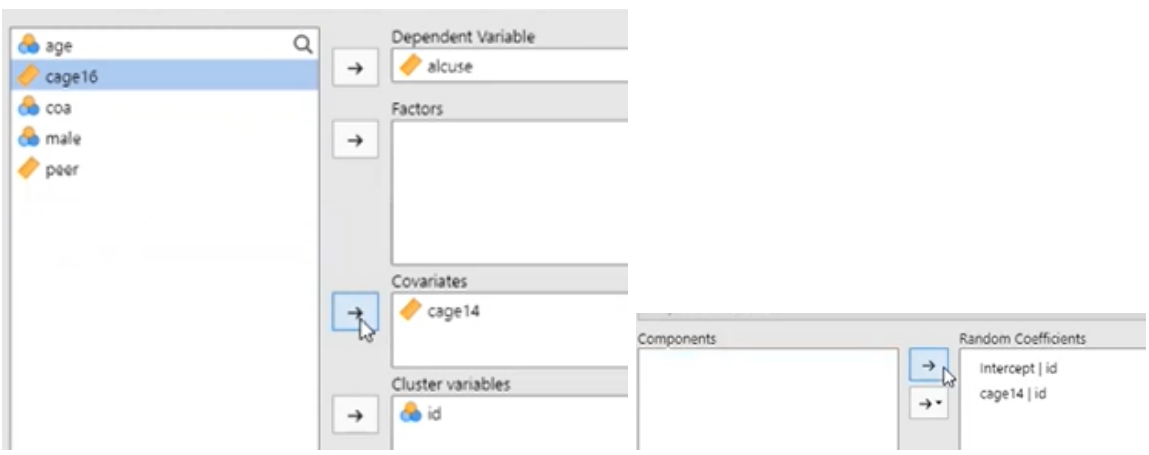
Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.757	0.573	0.505
Residual		0.749	0.562	

Note: Number of Obs: 246 , groups: id 82

%50 of variability in alcohol use is due to ind diff and the other % is due to aging (time) (how much of the variability in behav is explained by ind diff.(btw ind) And how much is explained by time(within ind))

Introduce time in the model:



We random our slope (time, age); so we allow the time slope to vary among Ps, because we are interested in ind variability (whether inds differ from each other in this average trajectories of development) as well as average patterns of development

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval Lower	Upper	df
(Intercept)	0.651	0.106	0.444	0.859	81.000
cage14	0.[illegible]	0.063	0.147	0.394	81.000

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.797	0.636	0.653
	cage14	0.394	0.155	
Residual		0.581	0.337	

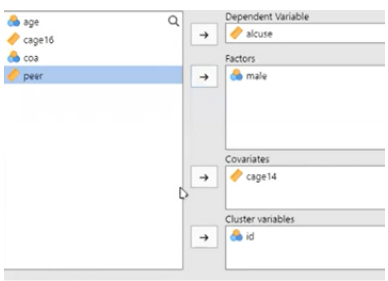
Note. Number of Obs: 246 , groups: id 82

Random Parameters correlations

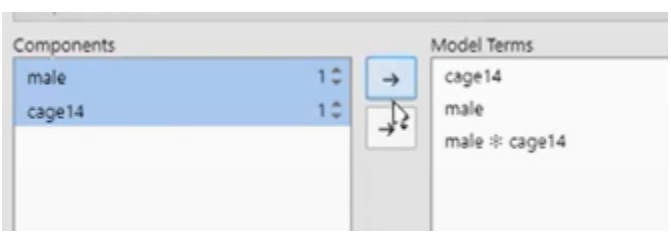
Groups	Param.1	Param.2	Corr.
id	(Intercept)	cage14	−0.227

A pos slope value (0.27) meaning that the mean amount of alcohol use increases linearly with age. As Ps get older they use more alcohol.

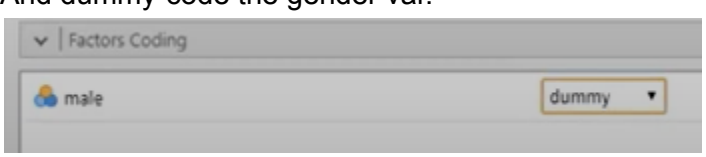
Whether this increase in alcohol use with age is steeper for some inds or not? Is there person-level predictors that effect the magnitude of the slope term?
→ add person-level predictors in the model: here we add gender variable



And we add the cross-level interaction term:



And dummy-code the gender var.



Fixed Effects Parameter Estimates

Name	Effect	Estimate	SE	95% Confidence Interval Lower	95% Confidence Interval Upper	df	t	p
(Intercept)	(Intercept)	0.746	0.152	0.448	1.043	80.000	4.918	< .001
cage14	cage14	0.121	0.087	−0.050	0.293	80.000	1.387	0.169
male1	1 - 0	−0.184	0.212	−0.599	0.231	80.000	−0.869	0.387
cage14 * male1	cage14 * 1 - 0	0.292	0.122	0.052	0.531	80.000	2.386	0.019

Random Components

Groups	Name	SD	Variance	ICC
id	(Intercept)	0.799	0.638	0.654
	cage14	0.371	0.137	
Residual		0.581	0.337	

Note. Number of Obs: 246 , groups: id 82

Random Parameters correlations

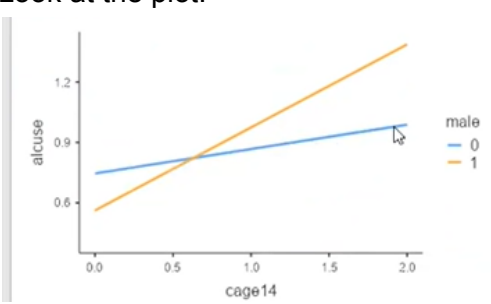
Groups	Param.1	Param.2	Corr.
id	(Intercept)	cage14	−0.205

Now, The intercept is mean alcohol use at age 14 among females (because females are dummy coded as 0).
And now, the slope is estimated rate of annual increase in alcohol use in females
So the slope (.12) gives me the rate of linear change with age in alcohol use in females

Looking at the interaction value (.292) → when I go from 0 to 1 (females to males), the rate of change is increased by .29 (difference btw males and females)

So, the rate of change in alcohol use differes across males and females (gender is a sig. Person-level predictor that effects the rate of change in alcohol use with age; males use more alcohol with age compared to females)

Look at the plot:



Rate of change for males is much more steeper

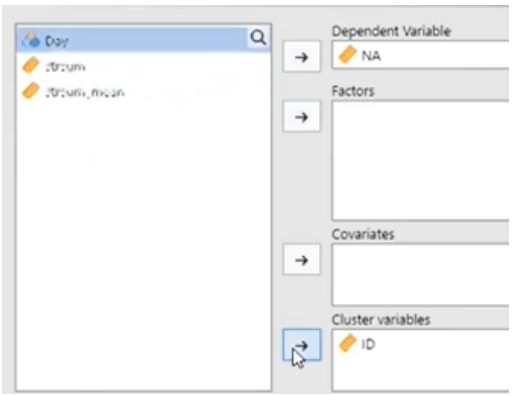
Growth Curve Models in Intensive Longitudinal data

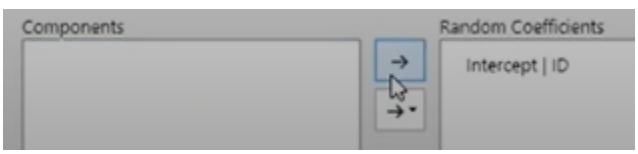
Jamovi exp: using level-1 parameter other than Time

Data: Ps assessed over 21 days, measures: negative affect and the num of stressors they experienced that day

Model: we are looking at the association between two level-1 parameters ; neg affect and stressors

→Start with a null model to see; what % of variability in NA in daily life at btw-Ps level and at within-Ps level (within person changes day to day)





Random intercept; we allow mean NA across 21 days for each Ps to vary randomly

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	2.551	0.067	2.420	2.682	150.006	38.137	< .001

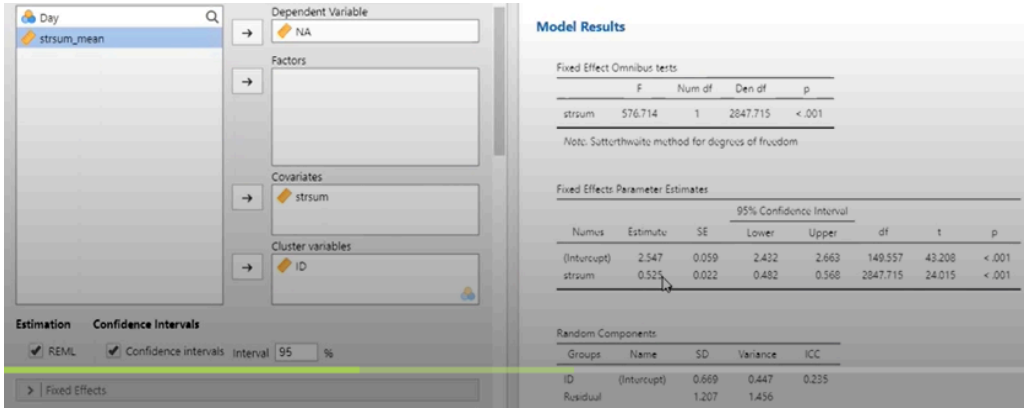
Random Components

Groups	Name	SD	Variance	ICC
ID	(Intercept)	0.763	0.582	0.249
Residual		1.324	1.753	

Note. Number of Obs: 2853 , groups: ID 151

%25 of variability in NA is due to difference btw ppl, and remaining variability in NA is due to changes within inds day to day

→ Add level-1 predictor (stressor):



Our intercept here (if we center our var. “stress”); is mean NA on average stressful days
(But if we do not center our stress var; then the intercept will tell you mean NA when there is 0 stress)

Slope value; 1 unit increase in num of stressors lead to .52 increase in NA

→ Repeated (within Ps) measures design; one advantage is assessing both btw-Ps and within-Ps differences.
****How can we disentangle btw-Ps effects from within-Ps effects?

Within-Ps question would be: Is there a diff btw stressful days and non-stressful days in terms of NA? (looking at the differences in NA across days) → looking at the unique effect of within-Ps change of stress on NA

Btw-Ps question would be: Are inds who are more stressed in general more likely to experience more NA? → looking at the unique effect of btw-Ps differences in stress on NA

Can I test these 2 questions simultaneously in this data? → YES!
In our exp data, we assessed both NA and stressor variables repeatedly over 21 days. So, we can test within-Ps and btw-Ps effects in the same model

**For every Ps, I have 21 days of data for stressor var.; get the average of these 21 observations → we would have mean stress exposure for every Ps, average for 21 days

ID	Day	NA	strsum	strsum_mean
1001	1	1.500	1	0.476
1001	2	4.000	3	0.476
1001	3	3.333	1	0.476
1001	4	1.000	2	0.476
1001	5	1.000	0	0.476
1001	6	1.500	0	0.476
1001	7	2.500	0	0.476
1001	8	2.000	0	0.476
1001	9	1.667	0	0.476
1001	10	2.333	2	0.476
1001	11	3.167	0	0.476
1001	12	2.167	0	0.476
1001	13	2.667	0	0.476
1001	14	3.833	0	0.476
1001	15	4.000	0	0.476
1001	16	4.000	0	0.476
1001	17	4.333	0	0.476
1001	18	3.667	0	0.476
1001	19	3.833	0	0.476
1001	20	3.667	1	0.476
1001	21	4.000	0	0.476
100	1	2.500	2	1.000

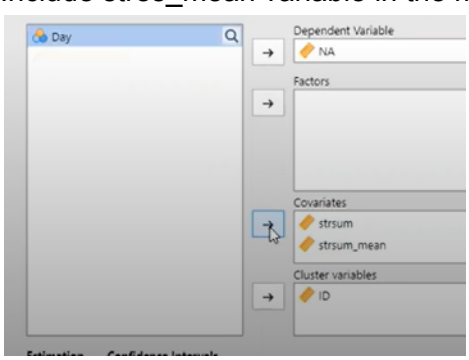
Thus, stres_mean in this dataset corresponds to an index of Ps' average stress in daily life.
We can use this to look at btw-Ps differences

And we also have daily stress fluctuations for each Ps across 21 days
We can use this to look at within-Ps differences

→ To model:
For within-Ps diff:

We use group/cluster based centering; meaning that I will subtract each person's mean score of stress from their daily stress score → the result will tell you the effect of stress on NA when you have more stressors than usual (compared to your own life, not compared to the overall grand mean/other Ps). Thus, 0 here is your own level of stress in daily life, we are putting 0 to your own average

For btw-ps diff:
Include stres_mean variable in the model



Fixed Effect Omnibus tests

	F	Num df	Den df	p
strsum	271.940	1	119.122	< .001
strsum_mean	43.525	1	149.183	< .001

Note: Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	2.547	0.059	2.432	2.663	148.586	43.192	< .001
strsum	0.546	0.033	0.481	0.611	119.122	16.491	< .001
strsum_mean	0.602	0.091	0.423	0.781	149.183	6.597	< .001

Random Components

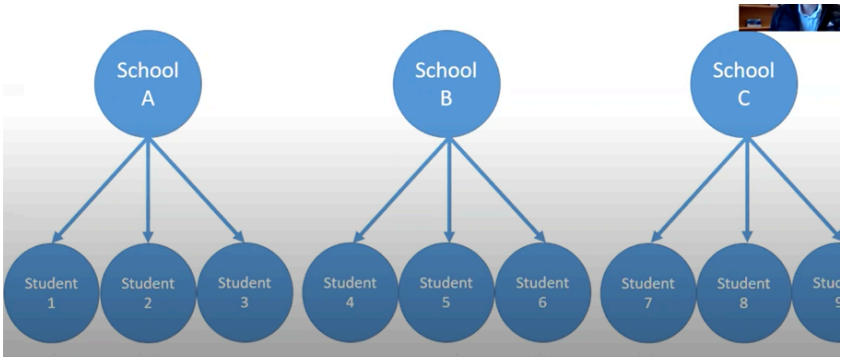
Groups	Name	SD	Variance	ICC
ID	(Intercept)	0.672	0.452	0.247
	strsum	0.278	0.077	

For within-ps level (.54); on days when ppl experience greater num of stressors than their usual level of stress, they experience more NA. (1 unit increase in stress results in .54 increase in NA)

For btw-ps level (.60); ind who in general have more stressful life, experienced greater NA

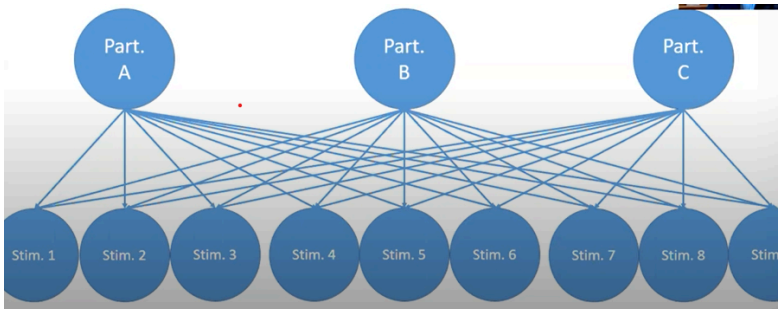
Cross-classified Models

So far, we have strict hierarchy in our mixed models; meaning that each student belonged only one and only school. The lower level units are mutually exclusive across clusters



That may not always be the case; sometimes you might identify random factors in your data and not see a strict hierarchy; exp; seeing a one random factor nested in multiple clusters
This is typical in cognitive psy research

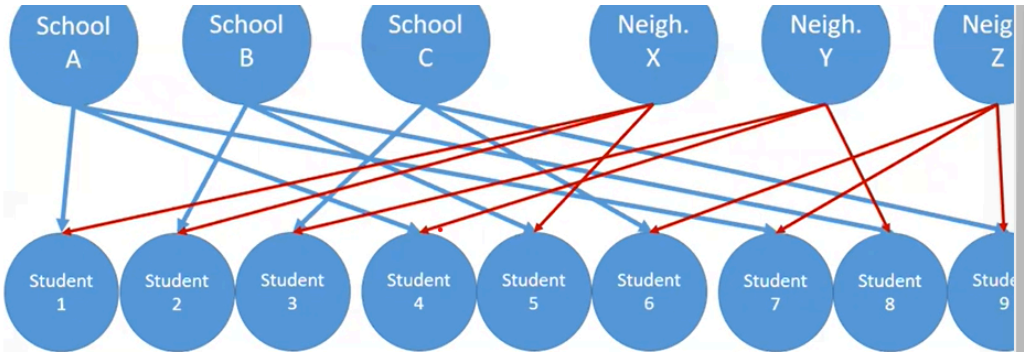
When all stimuli is presented to all Ps: (or, stimuli is evaluated by all Ps)



In this case there is no strict hierarchy but there are 2 random factors; Ps and stimuli

How can we decompose variability into these 2 sources of random variations? → cross classified models (there is no nesting structure but factors are crossed with each other)

Another exp:



Here, we cannot talk about neat structure btw school and negihbh. And students do not only belong to only 1 cluster; they belong to 1 school and 1 neighborh so the school and neighborh are crossed in a sense

Participants crossed with stimuli

Level 1

$$(\text{Trustworthiness})_{ijk} = \beta_{0jk} + e_{ijk}$$

Level 2

$$\beta_{0jk} = \gamma_{000} + u_{00j} + u_{00k}$$

Grand mean trustworthiness

Random effect of stimuli

Random effect of participants

When there is crossed factors; level 2 has more than 1 random effect in the null model

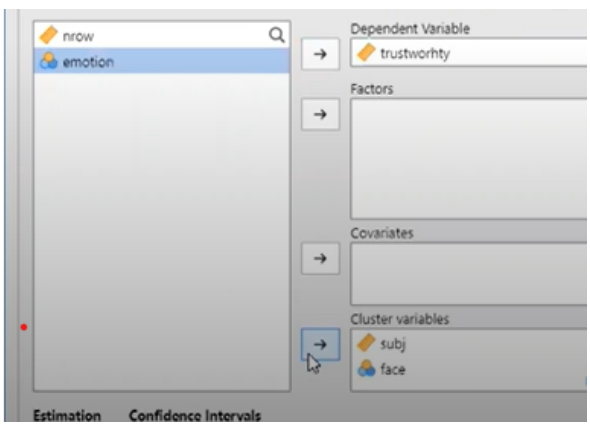
****the only difference of this modeling; when random factors are crossed (when there is no strict hierarchy btw random factors) you need to decompose variability into these crossed factors at lvl 2, by estimating more than one random effect

Jamovi exp:
Indicate trustworthiness judgment on various face stimuli, we have 30 different faces and 50 Ps; and each Ps see each face twice once happy, once angry
(whether emotion affects trustworthiness judgements?)

This data has not have a clear hierarchy; each face is evaluated by all Ps; so no nesting structure, there is cross-random factors

Here we have 2 random factors crossed with each other; Ps crossed with stimuli (there are no hierarchy or nested relationship btw them; rather a crossed structure → all Ps evaluate all stimuli and stimuli is evaluated by all Ps)

First; compose a null model to decompose variability into its sources;



(now we have 2 clustering variables)

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval Lower	Upper	df	t	p
(Intercept)	19.604	0.473	18.677	20.532	63.339	41.427	< .001

Random Components

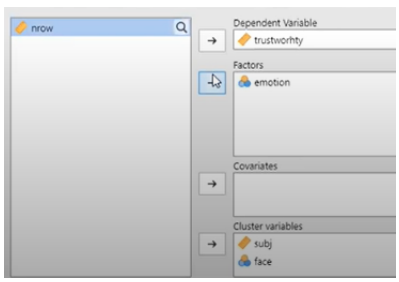
Groups	Name	SD	Variance	ICC
subj	(Intercept)	2.132	4.547	0.306
face	(Intercept)	1.972	3.887	0.274
Residual		3.208	10.293	

Note. Number of Obs: 3000 , groups: subj 50, face 30

The intercept (the grand mean) indicating the average trustworthiness

Different than before, We now estimated variability into 2 random factors at level 2 (subj and face)
We have 2 ICC now. → telling the amount of variability due to Ps (%30) and due to faces (%27)
(here, we reduce chance to make a type 1 error; because we identify random factors that explain a certain amount of variability in our data)

Then add predictors into the model (emotion, in this case):



Fixed Effects Parameter Estimates

Names	Effect	Estimate	SE	95% Confidence Interval Lower	Upper	df	t	p
(Intercept)	(Intercept)	19.119	0.477	18.185	20.054	65.295	40.096	< .001
emotion1	happy - angry	0.970	0.116	0.743	1.197	2920.000	8.377	< .001

Random Components

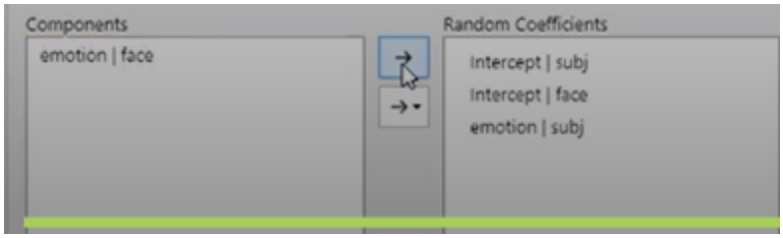
Groups	Name	SD	Variance	ICC
subj	(Intercept)	2.133	4.551	0.312
face	(Intercept)	1.972	3.890	0.279
Residual		3.171	10.055	

Note. Number of Obs: 3000 , groups: subj 50, face 30

Contrast Coefficients

There is an addition 1 point increase in trustworth score if the face is happy, a smiling face

Adding a random slope:
If you think that the effect of emotion on trustworthiness judgment may change person to person, you can randomize the slope **at the level of Ps**



Random Components

Groups	Name	SD	Variance	ICC
subj	(Intercept)	2.265	5.132	0.366
	emotion1	2.148	4.614	
face	(Intercept)	1.975	3.901	0.305
Residual		2.982	8.894	

Note. Number of Obs: 3000 , groups: subj 50, face 30

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
subj	(Intercept)	emotion1	−0.353

Here, now we have a random effect of emotion and a correlation btw slope (emotion) and intercept