



UNIVERSITAS AIRLANGGA



#SainsTerbuka  
**AIRLANGGA**  
Ilmu yang dibagi akan abadi



**IGDORE**  
Institute for Globally Distributed  
Open Research and Education

# Linear Mixed Models (Multi-Level Modeling)

## Menggunakan jamovi (GAMLj): Bagian 2

### Rizqy Amelia Zein

- Dosen, Fakultas Psikologi, Universitas Airlangga
- Anggota, #SainsTerbuka Airlangga 
- Relawan, INA-Rxiv
- Researcher-in-training, Institute for Globally Distributed Open Research and Education (IGDORE)

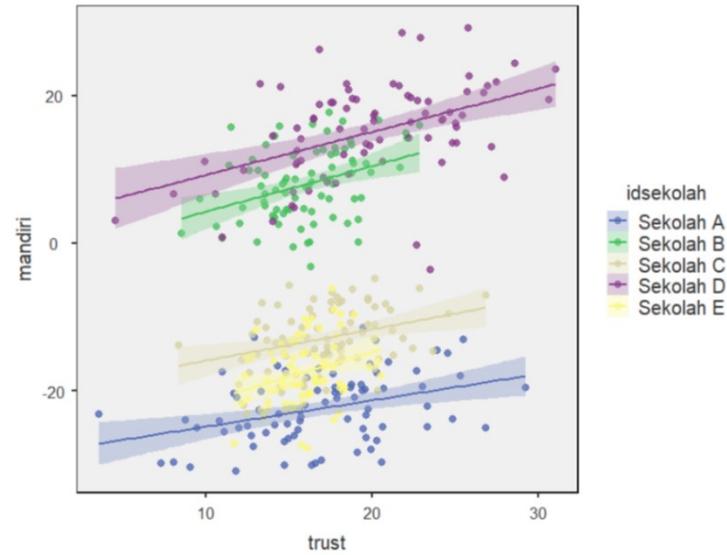
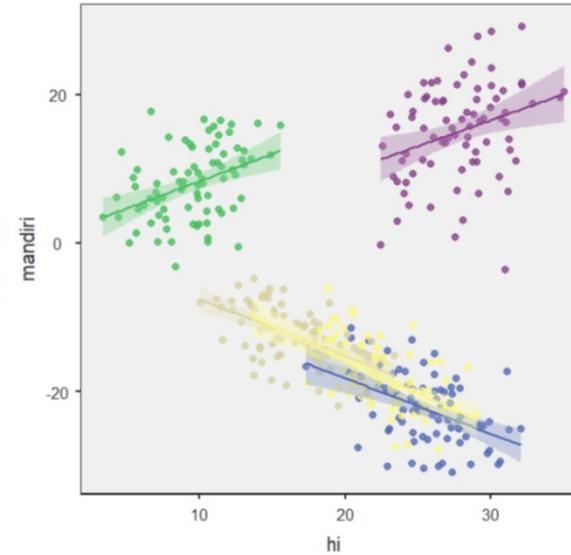
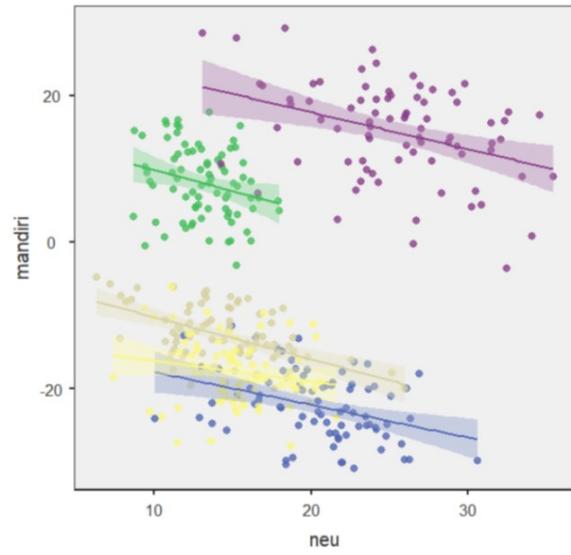
Quiz time! 😄

Klik disini untuk menuju laman kuis.



# Coba kita lihat lebih dekat...

- Coba kita lakukan inspeksi visual sekali lagi dataset-sekolah.omv
- Buat *scatterplot* dimana **mandiri** menjadi **Y-Axis**, sedangkan **neu, hi, trust** sebagai **X-Axis**
- Kemudian masukkan **idsekolah** pada kolom **Group**
  - Fungsinya, kita akan mendapatkan garis regresi untuk masing-masing sekolah
- Apa yang terjadi?

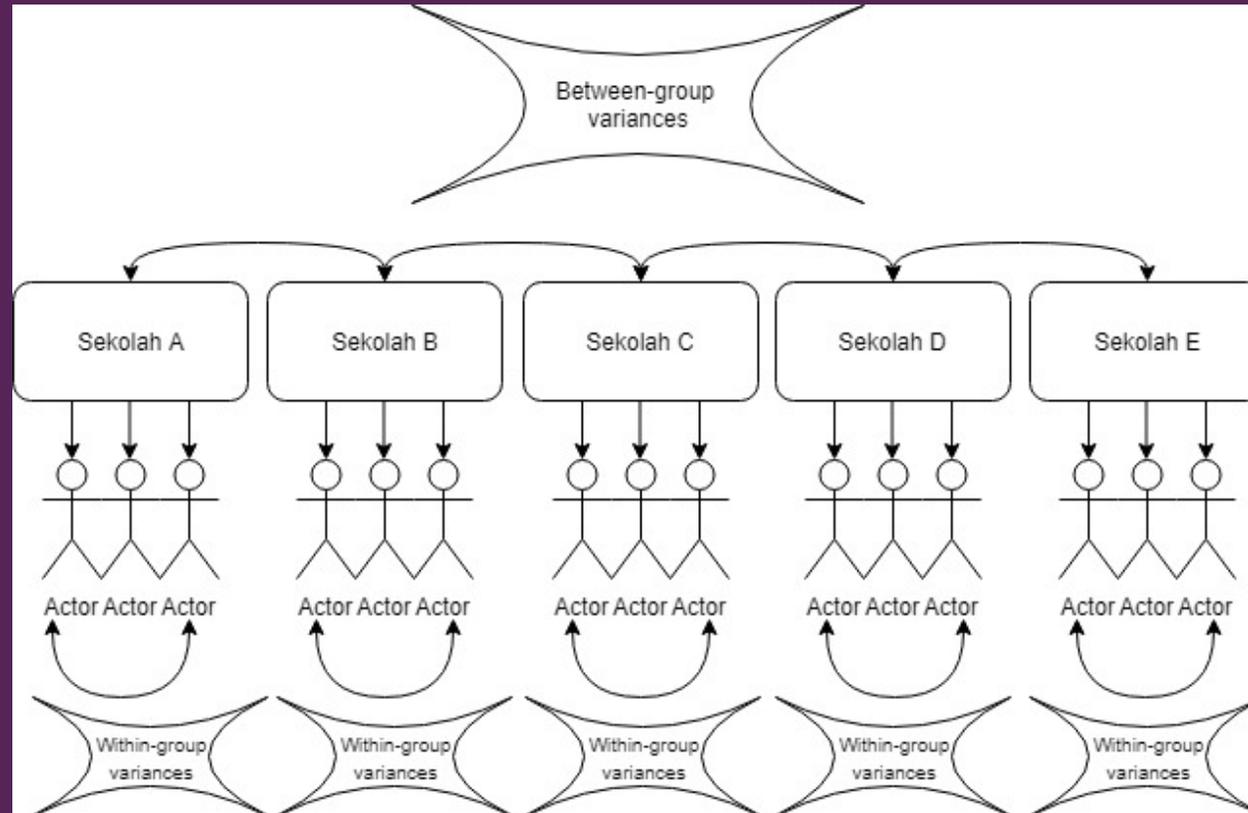


# Ternyata *covariancenya* bervariasi di tiap kelompok!

- *Intercept neuroticism* dan *trust* ternyata bervariasi di setiap sekolah
- Yang menarik, tidak hanya *intercept*, *slope* pendapatan personal juga bervariasi di setiap sekolah
  - Berdasarkan analisis yang kita lakukan pagi tadi, disimpulkan bahwa **pendapatan keluarga dan kemandirian anak korelasinya negatif**
  - Tapi bisakah **kesimpulan yang sama** kita tarik untuk Sekolah B dan D?
- Hati-hati *ecological fallacy*!
  - Terjadi ketika kita salah menyimpulkan suatu gejala yang skalanya individual, padahal yang dianalisis oleh peneliti sesungguhnya fenomena yang berlaku pada skala yang lebih besar (kelompok atau sub-kelompok)



# Struktur sampel bersarang/berjenjang



# Apa yang harus dilakukan?

- Pura-pura *ga tahu*, kita abaikan saja dan langsung menggunakan regresi OLS, dengan atau tanpa informasi mengenai pengelompokan data sebagai variabel kontrol.
  - Masalahnya, data/observasi kita sangat bergantung pada pengelompokan unit analisis
  - **Nah, lalu melanggar asumsi OLS dong?** (data/observasi dan residual harus independen)
  - Efeknya, *standard error* yang diestimasi oleh model terlalu kecil (karena mengabaikan varians dependen variabel yang ditentukan oleh kelompok)
  - Varians variabel dependen yang tidak bisa dijelaskan (residual) akan makin besar
  - Kesimpulan/inferensi yang ditarik menjadi tidak tepat, sehingga risiko terjadinya *type I error* meningkat.
- Gimana kalau pengelompokan (*group status*) dimasukkan aja dalam regresi OLS sebagai variabel moderator
  - Dengan begitu, estimasi *standard error* disesuaikan dengan menggunakan *marginal model*
  - Estimasi *standard error* akan lebih presisi, **tetapi** kita tetap tidak bisa mengestimasi *between-group variance*
- Kalo di agregat? Jadi unit analisis yang tadinya individual, menjadi kelompok.
  - Ukuran sampel menjadi lebih sedikit, sehingga *statistical power* menjadi lebih rendah!

# Fixed dan random effects

## Model fixed effects

$$y_{ij} = b_{0ij} + b_{1ij} * x_{ij} + e_{ij}$$

$$y_{i1} = b_{0i1} + b_{1i1} * x_{i1} + e_{i1}$$

$$y_{i2} = b_{0i2} + b_{1i2} * x_{i2} + e_{i2}$$

$$y_{i3} = b_{0i3} + b_{1i3} * x_{i3} + e_{i3}$$

$$y_{i4} = b_{0i4} + b_{1i4} * x_{i4} + e_{i4}$$

$$y_{i5} = b_{0i5} + b_{1i5} * x_{i5} + e_{i5}$$

Fixed intercepts =  $c_{00}$   
Agregat dari intercept semua level-2

Fixed slopes =  $c_{10}$   
Agregat dari slopes semua level-2

## Model random effects

$$u_{0j} = b_{0j} - c_{00}$$

$$u_{1j} = b_{1j} - c_{10}$$

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

Random intercepts =  $u_{0j}$   
Varians intercept di kelompok  $j$   
(level-2)

Random slopes =  $u_{1j}$   
Varians slopes di kelompok  $j$   
(level-2)

Level-1 residuals =  $e_{ij}$   
Varians residual (unexplained variance) antara orang  $i$  (level-1) di kelompok  $j$

# Full model

$$y_{ij} = b_{0j} + b_{1ij} * x_{ij} + e_{ij} \quad \text{Model level-1}$$

$$u_{0j} = b_{0j} - c_{00} \quad \text{Model level-2: random intercept}$$

$$u_{1j} = b_{1j} - c_{10} \quad \text{Model level-2: random slopes}$$

**Full model**

$$y_{ij} = (c_{00} + c_{10} * x_{ij}) + (u_{0j} + u_{1j} * x_{ij}) + e_{ij}$$

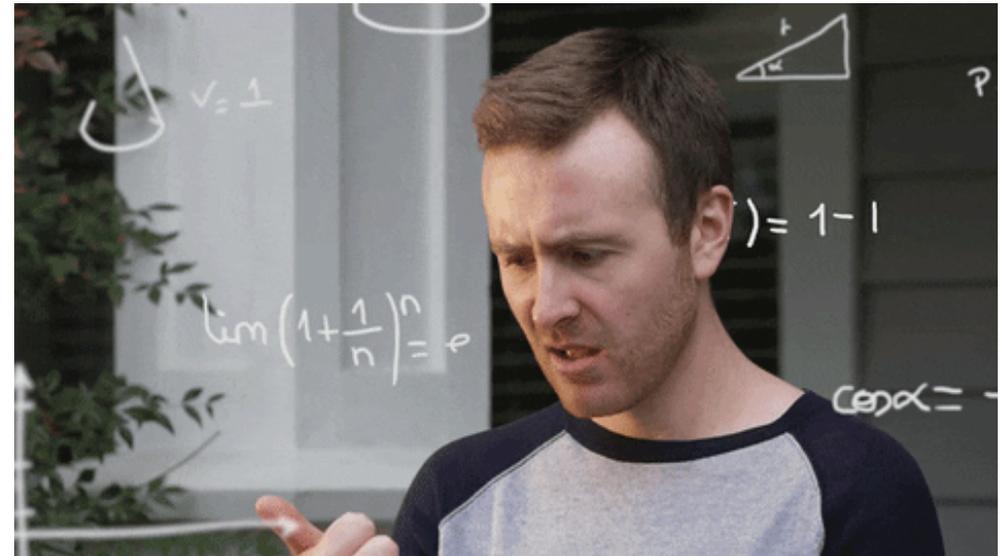
**Linear mixed effects**

**Fixed effects**

**Random effects**

# Kovarians (korelasi) antara *random intercept* dan *random slopes* ( $\sigma_{U_0U_1}$ )

- Nilainya **positif**, maka semakin tinggi *intercept* akan diasosiasikan dengan kemiringan garis yang lebih curam/*slopes* yang lebih besar
- Misalnya, di sekolah yang **rata-rata pendapatan** keluarga inti perbulan siswanya **tinggi**, maka **korelasi** antara pendapatan per bulan dengan tingkat kemandirian siswa **akan menguat**.
- Nilainya **negatif**, maka semakin tinggi *intercept* akan diasosiasikan dengan kemiringan garis yang lebih landai/*slopes* yang lebih kecil
- Misalnya, di sekolah yang **rata-rata pendapatan** keluarga inti perbulan siswanya **tinggi**, maka **korelasi** antara pendapatan per bulan dengan tingkat kemandirian siswa **akan melemah**.



# Parameter yang diestimasi dalam *linear mixed effects*

- *Fixed intercept* ( $c_{00}$ )
- *Fixed slopes* ( $c_{10}$ )
- *Varians random intercept* ( $\sigma^2_{U0}$ )
- *Varians random slopes* ( $\sigma^2_{U1}$ )
- *Kovarians (korelasi) antara random intercept dan random slopes* ( $\sigma_{U0U1}$ )
- *Varians residual level-1* ( $\sigma^2_e$ )

Yuk kita coba! 🖱️

Pastikan *module* GAMLj sudah terpasang di jamovi



## Latihan (3): Kembali ke sekolah

Setelah menginspeksi data secara visual, kita tahu bahwa korelasi antara **pendapatan keluarga** dengan **tingkat kemandirian anak** adalah yang paling bervariasi (daripada prediktor yang lainnya). Oleh karena itu, kita akan membuat *linear mixed model* dengan **pendapatan keluarga** sebagai prediktor, dan **tingkat kemandirian anak** sebagai variabel dependen.

### Bikin "model kosong"

- Yaitu model yang isinya hanya *intercept* saja, tidak ada prediktornya (*slopes*)
- Pada *menu bar*, klik **Linear Models**, pilih **mixed models**
  - Masukkan **mandiri** dalam kolom **dependent variable**
  - Masukkan **idsekolah** dalam kolom **cluster variables**
  - Pada menu **random effects** masukkan **intercept|idsekolah** dalam kolom **random coefficients**
- Catat nilai AIC yang tersedia dalam tabel **model info**
- Yuk ingat-ingat lagi, AIC fungsinya untuk apa ya? 😊

# Latihan (3)

## Bikin model dengan prediktor (*linear mixed model*)

- Masukkan **hi** dalam kolom **covariates**
- Pada menu **random effects**, masukkan juga **hi|idsekolah**, karena kita akan mengestimasi **random slopesnya** juga
  - Centang opsi **LRT for Random Test**
- Pada menu **covariates scaling**, ubah **centered** menjadi **cluster-based centered**
  - Berkaitan dengan **partitioning** (akan dijelaskan di bagian selanjutnya)

# Fixed coefficients

- Tes kecocokan model (*Omnibus Test*) tidak signifikan ( $F(1,3.98)=0.309, p=.608$ )
  - Berbeda sekali dengan hasil OLS
  - **Kemungkinan besar** korelasi antara pendapatan keluarga per bulan dengan kemandirian **tidak linier**
- Kita tidak punya cukup bukti untuk menolak hipotesis nol, bahwa pendapatan keluarga per bulan dengan kemandirian tidak berkaitan ( $B=-0.207$  95% CI [-0.937, 0.523],  $SE=0.372, t=-0.556, p=.608$ ).
  - Sekali lagi, berbeda sekali dengan hasil OLS

Fixed Effect Omnibus tests

	F	Num df	Den df	p
hi	0.309	1	3.98	0.608

Note. Satterthwaite method for degrees of freedom

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval		df	t	p
			Lower	Upper			
(Intercept)	-5.950	7.377	-20.409	8.508	4.00	-0.807	0.465
hi	-0.207	0.372	-0.937	0.523	3.98	-0.556	0.608

# Random coefficients

- Varians kemandirian paling banyak dijelaskan oleh rata-rata tingkat kemandirian anak di masing-masing sekolah (varians *random intercept* ( $\sigma^2_{U0}$ ) daripada oleh varians *random slopes* ( $\sigma^2_{U1}$ ).

Random Components

Groups	Name	SD	Variance	ICC
idsekolah	(Intercept)	16.488	271.848	0.931
	hi	0.814	0.663	
Residual		4.506	20.302	

Note. Number of Obs: 400 , groups: idsekolah , 5

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
idsekolah	(Intercept)	hi	0.983

Random Effect LRT

Test	N. par	AIC	LRT	df	p
hi in (1 + hi   idsekolah)	4	2454	74.0	2.00	8.567e-17

# Random coefficients

- Menguji efek sekolah (kelompok)
  - *Intra-class correlation*, yaitu merupakan proporsi total varians variabel dependen yang dapat dijelaskan oleh variasi antar kelompok
  - *Likelihood ratio test* (LRT), yaitu teknik untuk menguji ada/tidaknya perbedaan varians antar-kelompok
  - Keduanya juga bisa berfungsi sebagai indikator perlu/tidaknya `lme` dilakukan

Random Components

Groups	Name	SD	Variance	ICC
idsekolah	(Intercept)	16.488	271.848	0.931
	hi	0.814	0.663	
Residual		4.506	20.302	

Note. Number of Obs: 400 , groups: idsekolah , 5

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
idsekolah	(Intercept)	hi	0.983

Random Effect LRT

Test	N. par	AIC	LRT	df	p
hi in (1 + hi   idsekolah)	4	2454	74.0	2.00	8.567e-17

# Random coefficients

- Menguji efek sekolah (kelompok) / (*testing group effects*)
  - ICC=0.931, artinya 93.1% varians tingkat kemandirian anak dijelaskan oleh perbedaan kelompok. ICC diatas 0.1 biasanya menunjukkan *lme* adalah opsi yang lebih baik daripada OLS.
  - LRT juga menunjukkan bahwa kita dapat menolak hipotesis bahwa tidak adanya perbedaan varians tingkat kemandirian antar-kelompok (LRT(2)=74.0,  $p < .001$ ).

Random Components

Groups	Name	SD	Variance	ICC
idsekolah	(Intercept)	16.488	271.848	0.931
	hi	0.814	0.663	
Residual		4.506	20.302	

Note. Number of Obs: 400 , groups: idsekolah , 5

Random Parameters correlations

Groups	Param.1	Param.2	Corr.
idsekolah	(Intercept)	hi	0.983

Random Effect LRT

Test	N. par	AIC	LRT	df	p
hi in (1 + hi   idsekolah)	4	2454	74.0	2.00	8.567e-17

# Model Comparison

- AIC
  - Apabila kita membandingkan "model kosong" dengan model yang ada prediktor, maka model yang terakhir lebih mampu menjelaskan varians kemandirian anak.
- $R^2$  (Nakagawa & Schielzeth, 2012)
  - *Marginal*, merupakan proporsi varians variabel dependen yang dapat dijelaskan oleh **hanya fixed models** saja
  - *Conditional*, merupakan proporsi varians variabel dependen yang dapat dijelaskan oleh **fixed dan random models** sekaligus
  - Varians tingkat kemandirian anak yang dapat dijelaskan oleh *fixed model* saja hanya 0.12%, sedangkan yang dapat dijelaskan oleh keseluruhan model adalah 93.18%.

## Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	mandiri ~ 1 +( 1   idsekolah )
AIC	2452.081
R-squared Marginal	0.000
R-squared Conditional	0.915

## Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	mandiri ~ 1 + hi+( 1 + hi   idsekolah )
AIC	2372.03463
R-squared Marginal	0.00121
R-squared Conditional	0.93189

## Random Effect LRT

Test	N. par	AIC	LRT	df	p
(1   idsekolah)	2.00	3334	878	1.00	5.489e-193

## Random Effect LRT

Test	N. par	AIC	LRT	df	p
hi in (1 + hi   idsekolah)	4	2454	74.0	2.00	8.567e-17

# *Contextual effects & partitioning*

- *Within-group effect*
  - Seberapa besar selisih Y dari 2 orang yang berada di **kelompok yang sama**, ketika **selisih Xnya** sebesar 1 poin?
  - Seberapa besar perbedaan tingkat kemandirian dua orang anak yang berada dalam **sekolah yang sama**, ketika selisih **tingkat pendapatan keluarga** mereka berbeda sebesar 1 poin?
- *Between-group effect*
  - Seberapa besar selisih Y dari 2 orang yang berbeda (dari 2 kelompok yang berbeda), namun berada dalam **posisi relatif yang sama** (dibandingkan dengan rerata kelompok), ketika **selisih X** mereka sebesar 1 poin?
  - Seberapa besar perbedaan tingkat kemandirian 2 orang anak yang berasal dari **sekolah yang berbeda**, namun pada **posisi yang sama** apabila **dibandingkan dengan rata-rata** kemandirian anak-anak di **sekolah mereka masing-masing** (misal sama-sama yang paling/paling tidak manja), apabila **selisih pendapatan keluarga** mereka = 1 poin?
- *Contextual effect*
  - Seberapa besar selisih Y dua orang dari kelompok yang berbeda, namun dengan **X yang sama**, ketika **rerata X kelompoknya** berbeda sebesar 1 poin.
  - Seberapa besar selisih tingkat kemandirian 2 orang anak dari **sekolah yang berbeda**, namun memiliki **tingkat pendapatan keluarga yang sama** persis, ketika **rata-rata tingkat pendapatan keluarga anak-anak di sekolahnya** berbeda sebesar 1 poin?

# *Contextual effects & partitioning*

- Untuk menghitung *contextual effect*, kita harus melakukan *partitioning* terlebih dahulu
- Umumnya yang dipartisi/*centering* adalah **variabel X**, bukan Y
- *Group-mean centering*
  - Nilai X individu dikurangi rata-rata X kelompoknya
  - Pendapatan keluarga anak A dikurangi rata-rata pendapatan keluarga anak-anak di sekolahnya
- *Grand-mean centering*
  - Nilai X individu dikurangi rata-rata X pada seluruh sampel
  - Pendapatan keluarga anak A dikurangi rata-rata pendapatan keluarga seluruh anak yang menjadi sampel
- *Contextual effect = Between-group effect - Within-group effect*
  - *Contextual effect* yang positif artinya **kelompok** dengan **rata-rata X yang lebih tinggi**, cenderung memiliki *intercept* (rata-rata Y) yang **lebih tinggi**, atau begitu pula sebaliknya.
  - *Contextual effect* yang negatif artinya **kelompok** dengan **rata-rata X yang lebih tinggi**, cenderung memiliki *intercept* (rata-rata Y) yang **lebih rendah**, atau begitu pula sebaliknya.

## Latihan (4): *contextual effects*

- Lakukan `lme` dengan memasukkan `hi_group_centered` dan `hi_grand_mean_centered` dalam satu model yang sama
- Masukkan kedua variabel tersebut dalam **fixed coefficients** dan **random coefficients**
- Lihat *fixed slopes*-nya untuk kedua prediktor



# Contextual effects

- Kita tidak punya bukti yang meyakinkan bahwa kita dapat menolak hipotesis bahwa *within* ( $B=0.257$  95% CI [-0.115, 0.629],  $SE=0.190$ ,  $t=1.354$ ,  $p=.213$ ), maupun *between-group effect* ( $B=-0.454$  95% CI [-1.207, 0.299],  $SE=0.384$ ,  $t=-1.182$ ,  $p=.315$ ) tidak dapat menjelaskan varians tingkat kemandirian anak.
- *Contextual effects* = -0.711

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval		df	t	p
			Lower	Upper			
(Intercept)	-4.994	7.055	-18.822	8.835	3.95	-0.708	0.519
hi_group_centered	0.257	0.190	-0.115	0.629	7.96	1.354	0.213
hi_grand_mean_centered	-0.454	0.384	-1.207	0.299	3.32	-1.182	0.315

# Bagaimana melaporkannya? (1)

"...untuk menguji hipotesis bahwa ada perbedaan rerata tingkat kemandirian anak, dan korelasi antara pendapatan keluarga dengan tingkat kemandirian anak di masing-masing sekolah, peneliti melakukan analisis *linear mixed effect*.

Tingkat kemandirian anak dijelaskan sebagai fungsi dari tingkat pendapatan keluarga, dengan mengontrol asal sekolah (PAUD) anak. Sebelum melakukan analisis, tingkat pendapatan keluarga dipartisi dengan cara menguranginya dengan rata-rata tingkat pendapatan keluarga di masing-masing sekolah (*group-mean/cluster-based centering*).

Pengujian model menghasilkan kesimpulan bahwa model tidak cocok menggambarkan data ( $F(1,3.98)=0.309, p=.608$ ), sehingga kami gagal menolak hipotesis bahwa tingkat pendapatan keluarga dan kemandirian anak, tidak berkorelasi. Namun, ada kemungkinan korelasi antara tingkat pendapatan keluarga dan kemandirian anak tidak linier.

Model *fixed effects* menunjukkan bahwa peneliti tidak punya cukup bukti untuk menolak hipotesis nol, bahwa pendapatan keluarga per bulan dengan kemandirian tidak berkaitan ( $B=-0.207$  95% CI  $[-0.937, 0.523]$ ,  $SE=0.372$ ,  $t=-0.556, p=.608$ ).

## Bagaimana melaporkannya? (2)

"...model *random effects* menunjukkan bahwa peneliti dapat menolak hipotesis bahwa tidak ada perbedaan varians tingkat kemandirian antar-kelompok (LRT(2)=74.0,  $p < .001$ ). Varians kemandirian anak paling banyak dijelaskan oleh rata-rata tingkat kemandirian anak di masing-masing sekolah (varians *random intercept* ( $\sigma^2_{U0}$ )) daripada oleh varians *random slopes* ( $\sigma^2_{U1}$ ). Selain itu, 93.1% varians tingkat kemandirian anak dijelaskan oleh perbedaan kelompok (ICC=0.931).

Peneliti tidak memiliki cukup bukti yang meyakinkan untuk menolak hipotesis yang menyatakan bahwa *within* ( $B=0.257$  95% CI [-0.115, 0.629],  $SE=0.190$ ,  $t=1.354$ ,  $p=.213$ ), maupun *between-group effect* ( $B=-0.454$  95% CI [-1.207, 0.299],  $SE=0.384$ ,  $t=-1.182$ ,  $p=.315$ ) tidak dapat menjelaskan varians tingkat kemandirian anak.

*Contextual effects* ditemukan sebesar -0.711, artinya, anak yang bersekolah di dua tempat yang berbeda, dengan selisih rata-rata tingkat kemandirian anak-anak di dua sekolah tersebut sebesar 1 poin, maka tingkat kemandirian mereka berbeda sebesar -0.711 poin, apabila diasumsikan keluarga mereka memiliki pendapatan yang sama besarnya. Tanda negatif dari *contextual effect* mengindikasikan bahwa sekolah yang rata-rata pendapatan keluarga siswanya lebih tinggi, cenderung memiliki rata-rata tingkat kemandirian yang rendah..."

# Latihan mandiri (2)

- Lakukan analisis `lme` untuk mengetahui:
  - Apakah varians **tingkat kemandirian anak** dapat dijelaskan oleh sekolah tempat anak tersebut belajar?
  - Apakah varians korelasi antara kecenderungan **neuroticism** ibu dengan **kemandirian anak** juga dapat dijelaskan oleh sekolah tempat anak tersebut belajar?
  - Seberapa besar perbedaan **tingkat kemandirian** dua orang anak yang berada di **sekolah yang berbeda**, yang **ibunya sama-sama pencemas**, apabila **rata-rata kecemasan wali murid di dua sekolah tersebut** berbeda sebesar 1 poin?

Klik disini untuk mengakses lembar kerja

# Yang belum dibahas...

- Kalau korelasi antara X dan Y tidak linier, pakai apa dong?
  - Jelas tidak bisa menggunakan *lme*. Alternatifnya, bisa menggunakan *generalised additive model (GAM)*.
- Kalau prediktornya level-2, bagaimana?
- Bisa ga *lme* digunakan untuk mengestimasi perubahan Y pada *time-series*?
- Gimana cara merencanakan jumlah sampelnya?
- Gimana kalo sampelnya bersarang/berjenjang level-3, bahkan lebih?
- Gimana kalo terjadi interaksi antara variabel prediktor level-1 dengan level-2 (*cross-level interactions*)?



# The problem with linear relationship



Terima kasih banyak! 😊



Paparan disusun dengan menggunakan [R package xaringan](#) dengan *template* dan *fonts* dari [R-Ladies](#).

*Chakra* dibuat dengan [remark.js](#), [knitr](#), dan [R Markdown](#).