Multilevel analysis

Three level models, assumptions and some other statistical notes

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

1

Last Monday

1) Introduction: why use multilevel analysis?
2) Building the multilevel regression model
3) Analysis approach
4) Example

2

Multilevel regression

At the lowest (individual) level we have

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

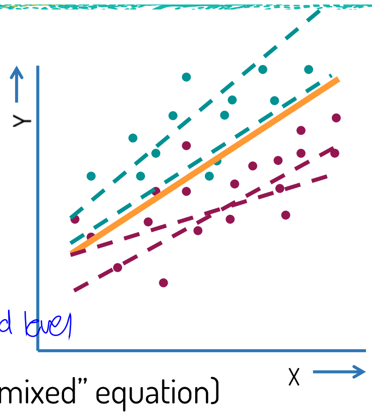
cluster-specific intercept

cluster-specific slope

and at the second (cluster) level.

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

can be seen as error terms in 2nd level



Combining (substitution and rearranging terms) gives ("mixed" equation)

$$y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij} + u_{0j} + u_{1j}X_{ij} + e_{ij}$$

Multilevel regression – interpretation

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

- the fixed part is an ordinary regression model
- complicated error term: $[u_{0j} + u_{1j}X_{ij} + e_{ij}]$

- Several error (co-)variances
  - σ_e^2 variance of the lowest level errors e_{ij}
  - σ_{u0}^2 variance of the highest level errors u_{0j}
  - σ_{u1}^2 variance of the highest level errors u_{1j}
  - σ_{u01} covariance of u_{0j} and u_{1j}

intercept var. & slope var.

- *Intraclass correlation* ρ:
  - $\rho = \sigma_{u0}^2/(\sigma_e^2 + \sigma_{u0}^2)$

quantifies dependency we have in our data: 0 → no dependency; 1 → complete dependency

Today

1) Three level data
2) Assumptions in linear multilevel models
3) Statistical notes

Three level data

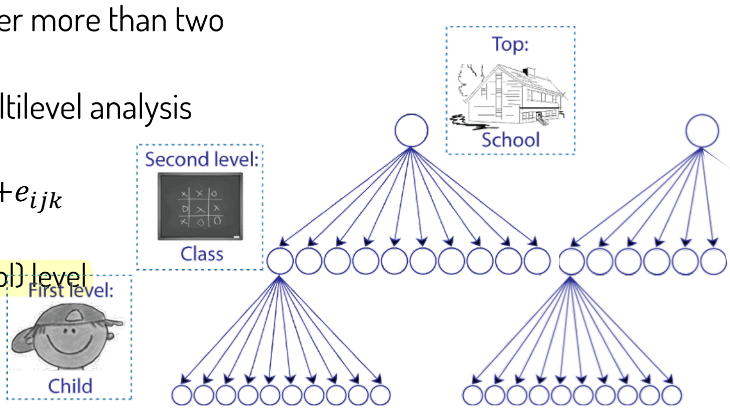
Pupils nested in classes nested in schools

Three level data

- In nested data, we may encounter more than two levels (e.g., three levels)
- Possible to accommodate in multilevel analysis
  - 'Static' data (mixed eq):

    $y_{\cancel{tik}}^{ijk} = \gamma_{000} + v_{0k} + u_{0jk} + e_{ijk}$
  - γ_{000} overall intercept
  - v_{0k} variance at third (e.g. school) level
  - u_{0ij} variance at second level
  - e_{ijk} residual error term

person-specific deviation to overall intercept



longitudinal data) mixed equation

$y_{tik} = \beta_{000} + v_{0k} + u_{0ik} + e_{tik}$

Three level data - ICC

- $\sigma_{v_0}^2$ variance of the third level errors v_{0k}
- $\sigma_{u_0}^2$ variance of the second level errors u_{0ik}
- σ_e^2 variance of the residual error term e_{tik}

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

$$\rho_{class} = \frac{\sigma_{u_0}^2}{\sigma_{v_0}^2 + \sigma_{u_0}^2 + \sigma_e^2} \qquad \rho_{school} = \frac{\sigma_{v_0}^2}{\sigma_{v_0}^2 + \sigma_{u_0}^2 + \sigma_e^2}$$

the way you compute, depends on what you wanna represent
do you wanna decompose the variance over diff. levels
OR
do you wanna calculate the exp. corr. between two observed sample from the same level

07/02/2022

100%

If we have

Three level data - ICC

there're multiple ways to compute ICC. (at the intermediate level)

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

by default, they come from the same school & therefore we need to take the var. at school level into account here.

- DIY: What about the interpretation 'expected correlation between two randomly chosen persons from the same class'? Do the equations have to be adjusted? How?

if we're looking at the expected corr., we need to account for the fact that the persons originate from the same class, & by default, they come from the same school. So both of'em need to be included in the denominator, if that's the interpretation that we're looking for.

$$\left(\frac{\sigma^2_{v_0} + \sigma^2_{u_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e} \right)$$

Assumptions

And how to check them

Multilevel regression – assumptions

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

- Individual level errors e_{ij} independent; normal distribution with mean zero and same variance σ_e^2 in all groups homoscedasticity
- Cluster level errors u_{0j} and u_{1j} are independent of e_{ij}
- Cluster level errors u_{0j} and u_{1j} have a multivariate normal distribution with means 0 and (co)variances $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01}$
- Plus usual assumptions of multiple regression analysis:
  - linear relations
  - no outliers
  - explanatory variables measured without error
  - absence multicollinearity

11

Multilevel regression – assumptions

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

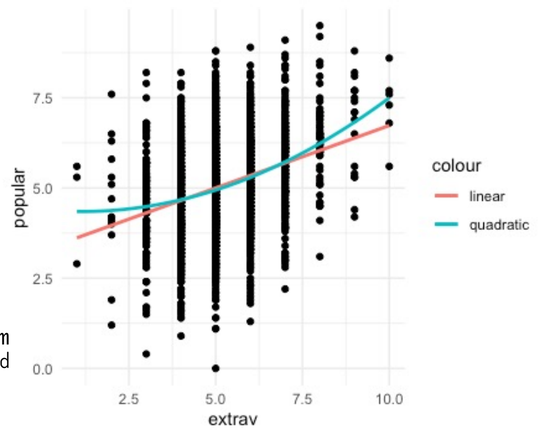
- Individual level errors e_{ij} independent; normal distribution with mean zero and same variance σ_e^2 in all groups
- Cluster level errors u_{0j} and u_{1j} are independent of e_{ij}
- Cluster level errors u_{0j} and u_{1j} have a multivariate normal distribution with means 0 and (co)variances $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01}$
- Plus usual assumptions of multiple regression analysis:
  - linear relations
  - no outliers
  - explanatory variables measured without error
  - absence multicollinearity

12

Before fitting ML model – linear relations and outliers

→ You can just use scatterplot for this

```
ggplot(popular,
aes(x = extrav, y = popular)) +
  geom_point() +
  geom_smooth(method = "lm",
              aes(color = "linear"),
              se = FALSE) +
  geom_smooth(method = "lm",
              formula = y ~ x + I(x^2),
              aes(color = "quadratic"),
              se = FALSE) +
  theme_minimal()
```



Notes:

Do not use `abline()`, this will only take the first two values of your `lm` object (so the intercept and slope), and will ignore the quadratic trend and not warn you about this.

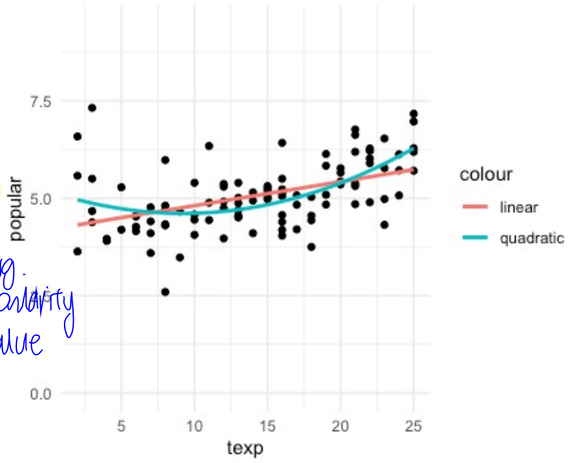
If you don't want to use ggplot, use something like `plot(x, y)` with `lines(fitted(…))` instead.

Before fitting ML model – linear relations and outliers

```
popular_aggr <- aggregate(popular, list(class
= popular$class), mean)
```

- For level 2 predictors, this means the dependent variable needs to be aggregated to the cluster means
- Same holds for outliers

avg. popularity value



Multilevel regression – assumptions

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

- Individual level errors e_{ij} independent; normal distribution with mean zero and same variance σ_e^2 in all groups
  but they're related to each other
- Cluster level errors u_{0j} and u_{1j} are independent of e_{ij}!
- Cluster level errors u_{0j} and u_{1j} have a multivariate normal distribution with means 0 and (co)variances $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01}$
- Plus usual assumptions of multiple regression analysis:
  - linear relations
  - no outliers
  - explanatory variables measured without error
  - absence multicollinearity

15

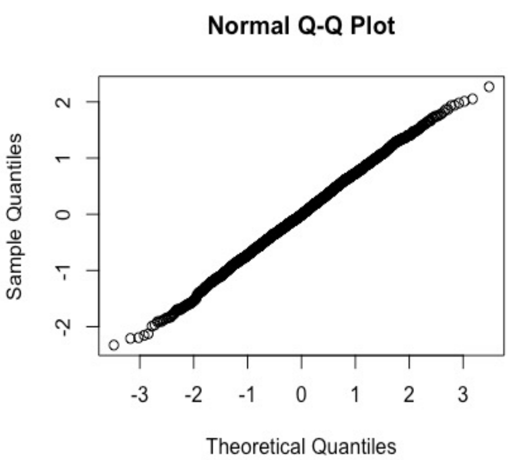
(error terms come after you fit a model)

After fitting final ML model – assumptions on residuals 1

```
model_5 <- lmer(popular ~ 1 + extrav + gender +
                texp + extrav*texp +
                (extrav|class),
                REML = FALSE, data = popular)

qqnorm(residuals(model_5))
```

- Use `qqnorm()` on object containing the final ML model for level 1 residuals
- Heteroscedasticity can be checked by plotting the predicted values against the residuals.



16

Multilevel regression – assumptions

$$y_{ij} = [\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{ij}] + [u_{0j} + u_{1j}X_{ij} + e_{ij}]$$

- Individual level errors e_{ij} independent; normal distribution with mean zero and same variance σ_e^2 in all groups
- Cluster level errors u_{0j} and u_{1j} are independent of e_{ij}
- Cluster level errors u_{0j} and u_{1j} have a multivariate normal distribution with means 0 and (co)variances $\sigma_{u0}^2, \sigma_{u1}^2, \sigma_{u01}$
- Plus usual assumptions of multiple regression analysis:
  - linear relations
  - no outliers
  - explanatory variables measured without error
  - absence multicollinearity

17

you also need to check for residuals at level 2.

After fitting final ML model - assumptions on residuals 2

```
# intercept variance
qqnorm(ranef(model_5)$class[,1])

# slope variance
qqnorm(ranef(model_5)$class[,2])
```

- Extract the cluster specific deviances to the intercept and slope using `ranef()`
- Use `qqnorm()` on the resulting object

random effects

```
> ranef(model_5)
$class
  (Intercept)        extrav
1 -0.902209893  3.829956e-02
2 -0.665647096  7.600675e-03
3 -0.542499173  3.750422e-02
4  0.092561952  1.062262e-02
5  0.602455419  2.033939e-02
6  0.103241470 -4.952523e-03
7 -0.726239374 -1.798017e-02
8 -1.181153940  4.369683e-02
```

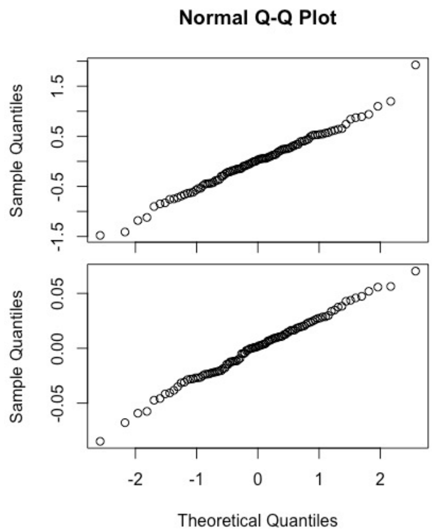

18

After fitting final ML model - assumptions on residuals 2

```
# intercept variance
qqnorm(ranef(model_5)$class[,1])

# slope variance
qqnorm(ranef(model_5)$class[,2])
```

- Extract the cluster specific deviances to the intercept and slope using `ranef()`
- Use `qqnorm()` on the resulting object



19

(Statistical) notes

Interaction effects, interpretation regression coefficients, p-values

20

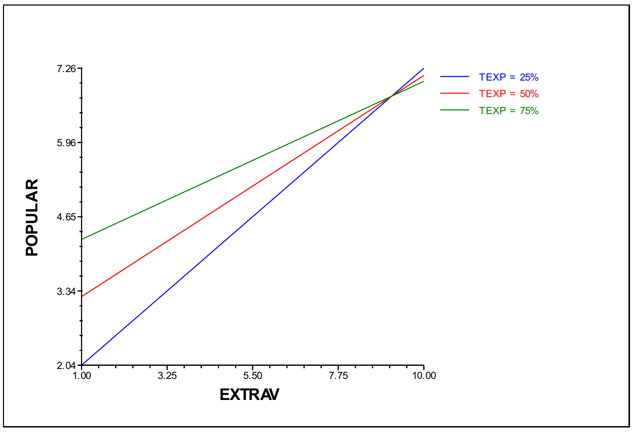
Note on Interpretation of Interaction

- The regression coefficients are *unstandardized* coefficients
- The regression coefficients change a lot when the cross-level interaction is added to the model becuz in the final model you gonna explain slope variance w/ interaction term.
- This is expected: in a model with interactions the regression coefficients for the direct effects are the expected value when the other predictor has value zero,
- which value does not even exist for *extraversion*
- Solution (1): plot the interaction
- Solution (2): *center* the predictor variables on their grand mean before computing interaction

Emmeke Aarts | Multilevel analysis - lecture 1b | 21

21

Note on Interpretation of Interaction



Emmeke Aarts | Multilevel analysis - lecture 1b | 22

22

Note on the Regression Coefficients

The regression coefficients are *unstandardized* coefficients

- Scale of predictor important for interpretation
  - sex 0-1,
  - extraversion 1-10,
  - teacher experience in years (2-25)

→ so we cannot just compare the magnitude of the predictors to see which has the largest effects on outcomes.

- Interpretation often based on *standardized* coefficients:

$$\gamma_{standardized} = (\gamma_{unstandardized} * s_x)/s_y$$

- Almost none of the present software produces standardized regression coefficients

23

Note on the Regression Coefficients

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

Unstandardized estimates (SE)	Standardized estimates
$\gamma_{00} = 0.74\ (.20)$	-
$\gamma_{10} = 1.25\ (.04)$	0.46
$\gamma_{20} = 0.45\ (.02)$	0.41
$\gamma_{01} = 0.09\ (.01)$	0.43

Teacher experience now appears more important

24

Note on *p*-values

- P values for regression coefficients: Wald test anova() function
- P values of variance components: do not use Wald test but chi-square model test
  - For two nested models, under H0 (no difference), the difference in deviance follows chi-squared squared distribution with df equal to difference in number of parameters
  - Only when using full maximum likelihood estimation
- P values of variance components: always one sided !! ∵ variance can't be neg.
  - Hence, for Wald or deviance tests of variances, take p/2.

Note on centering and slope variance

- If the model contains slope variance, the variance of the intercept depends on how the *X*-variable is centered.
- Also, depending on the centering, the covariance between the intercept and the slope can be positive as well as negative.
- We do not want that the fit of our model is influenced by this and therefore, when including slope variance, we also include covariance parameter!