# Multilevel Modelling: Binary Logistic Regression

Violetta Korsunova
Dep. of Sociology, HSE
March 17, 2020

# Why use multilevel logistic approach?

Like you may guess:

- Your data is hierarchical
- Your dependent variable is binary (it only has two possible outcomes)

# Multilevel Binary Logistic Regression

- Like in linear multilevel models, you assume that a part of your dependent variable's variance is determined by the differences in contextual factors aka 2-level factors

- As the outcome is non-continuous, you move to modelling the log of odds ratio

# Random intercept model

The common equation for a linear model goes as follows:

$$log\frac{P(yij = 1)}{1 - P(yij = 1)} = b_0 + b_1 * xi_j + ei_j$$

Random intercept model implies a unique b0 in each group:

$$b_{0j} = \gamma_{00} + \eta_{0j}$$

where $\gamma_{00}$ is the mean value of intercepts aka grand mean, and $\eta_{0j}$ is the error term that discerns the intercept in group j from the grand mean

The final equation is: $log\frac{P(yij=1)}{1-P(yij=1)} = \gamma_{00} + \eta_{0j} + b_1 * xi_j + ei_j$

# Random slope model

Again, the common equation for a linear model goes as follows:

$$log\frac{P(yij = 1)}{1 - P(yij = 1)} = b_0 + b_1 * xi_j + ei_j$$

Random slope model implies a unique b1 in each group:

$$b_{1j} = \gamma_{10} + \eta_{1j}$$

where $\gamma_{10}$ is the mean value of slopes aka grand slope, and $\eta_{1j}$ is the error term that discerns the slope in group j from the grand slope

The final equation is: $log\frac{P(yij=1)}{1-P(yij=1)} = b_0 + (\gamma_{10} + \eta_{1j}) * xi_j + ei_j$

# Random slope and intercept model

Finally, both intercepts and slopes can vary:

$$b_{0j} = \gamma_{00} + \eta_{0j}$$
$$b_{1j} = \gamma_{10} + \eta_{1j}$$

The full model is:

$$log\,\frac{P(yij = 1)}{1 - P(yij = 1)} = \gamma_{00} + \eta_{0j} + (\gamma_{10} + \eta_{1j}) * xij + ei_j$$

or

$$log\,\frac{P(yij = 1)}{1 - P(yij = 1)} = \gamma_{00} + \eta_{0j} + \gamma_{10} * x_{ij} + \eta_{1j} * xi_j + ei_j$$

# Inserting 2-level variable: random intercept

These models imply that a part of the group-level error terms (η) is explained by some factors (Z)

$$b_{0j} = \gamma_{00} + \eta_{0j}$$

$$\eta_{0j} = \gamma_{01} * Zj + e_{0j}$$

So the full equation:

$$log\frac{P(yij = 1)}{1 - P(yij = 1)} = \gamma_{00} + \gamma_{01} * Zj + e_{0j} + b_1 * xi_j + ei_j$$

# Inserting 2-level variable: random slope

$$b_{1j} = \gamma_{10} + \eta_{1j}$$
$$\eta_{1\mathrm{j}} = \gamma_{11} * Z_j + e_{1j}$$

So the full equation:

$$log \frac{P(yij = 1)}{1 - P(yij = 1)} = \gamma_{00} + \eta_{0j} + \gamma_{10} * x_{ij} + \gamma_{11} * Z_j * xi_j + e_{1j} * xi_j + ei_j$$

Again, the $\gamma_{11} * Z_j * xi_j$ term is called ***cross-level interaction***

# Inserting 2-level variable: random slope and intercept

As we combine both parts we can get the final model:

$$b_{0j} = \gamma_{00} + \eta_{0j}$$
$$\eta_{0\mathrm{j}} = \gamma_{01} * Z_j + e_{0j}$$
$$b_{1j} = \gamma_{10} + \eta_{1j}$$
$$\eta_{1\mathrm{j}} = \gamma_{11} * Z_j + e_{1j}$$

So the full equation:

$$log\frac{P(yij = 1)}{1 - P(yij = 1)} = \gamma_{00} + \gamma_{01} * Z_j + e_{0j} + \gamma_{10} * x_{ij} + \gamma_{11} * Z_j * xi_j + e_{1j} * xi_j + ei_j$$

# Intraclass correlation coefficient (ICC)

The issue with ICC in MBLR is that you don’t have real variance in the outcome

The value of individual variance is set to be always $\boldsymbol{\frac{\pi^2}{3}}$

$$ICC = \frac{\sigma^2(group)}{\sigma^2(group) + \frac{\pi^2}{3}}$$

Like in linear models, ICC should be at least 0.05 which means that 5% of the variance is due to group features

# How to do in R

Let’s look at the data ‘Europe2.RData’:

- cultinc – use of internet for cultural purposes (0 – no, 1 – yes)
- eduscaled – scaled age when completed education
- agescaled – scaled age
- gender – gender of the respondent (male, female)
- bill – if the respondent had difficulties paying bills (0 - yes, 1 – no)
- loggdp – log of GDP per capita
- country – country

Let’s predict the use of internet for cultural purposes

# Fit the null model

```
> library(lme4)
> m0 = glmer(cultinc~1|country, data = cult1, family = binomial)
> summary(m0)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
 Family: binomial  ( logit )
Formula: cultinc ~ 1 | country
   Data: cult1
     AIC      BIC   logLik deviance df.resid
 36604.5  36620.9 -18300.2  36600.5    27561
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7697 -1.0077  0.5658  0.8731  1.4115
Random effects:
 Groups  Name        Variance Std.Dev.
 country (Intercept) 0.2543   0.5042
Number of obs: 27563, groups:  country, 30
```

# Fit the null model

```
> library(lme4)
> m0 = glmer(cultinc~1|country, data = cult1, family = binomial)
> summary(m0)

Generaliz                  model fit by maximum likelihood (Laplace Approximation)
['glmer'
 Family:                          )
Formula: cultinc ~ 1 | country
   Data: cult1
     AIC       BIC    logLik deviance df.resid
 36604.5   36620.9 -18300.2  36600.5    27561
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7697 -1.0077  0.5658  0.8731  1.4115
Random effects:
 Groups  Name        Variance Std.Dev.
 country (Intercept) 0.2543   0.5042
Number of obs: 27563, groups:  country, 30
```

glmer
function

# Fit the null model

```
> library(lme4)
> m0 = glmer(cultinc~1|country, data = cult1, family = binomial)
> summary(m0)

Generalized linear mixed model fit by maximu                    ce Approximation)
['glmerMod']
 Family: binomial  ( logit )
Formula: cultinc ~ 1 | country
   Data: cult1
     AIC      BIC   logLik deviance df.resid
 36604.5  36620.9 -18300.2  36600.5    27561
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.7697 -1.0077  0.5658  0.8731  1.4115
Random effects:
 Groups  Name        Variance Std.Dev.
 country (Intercept) 0.2543   0.5042
Number of obs: 27563, groups:  country, 30
```

set the family

# Fit the null model

```
> library(lme4)
> m0 = glmer(cultinc~1|country, data = cult1, family = binomial)
> summary(m0)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
 Family: binomial  ( logit )
Formula: cultinc ~ 1 | country
   Data: cult1
     AIC      BIC   logLik deviance df.resid
 36604.5  36620.9 -18300.2  36600.5    27561
Scaled residuals:
    Min      1Q  Medi                  Max
-1.7697 -1.0077  0.5                  4115
Random effects:
 Groups  Name        Variance Std.Dev.
 country (Intercept) 0.2543   0.5042
Number of obs: 27563, groups:  country, 30
```

Group variance

# Fit the null model

```
> library(lme4)
> m0 = glmer(cultinc~1|country, data = cult1, family = binomial)
> summary(m0)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']
 Family: binomial  ( logit )
Formula: cultinc ~ 1 | country
   Data: cult1
     AIC      BIC   logLik deviance df.resid
 36604.5  36620.9 -18300.2  36600.5    27561
Scaled residuals:
    Min      1Q  Medi                    Max
-1.7697 -1.0077  0.5                  4115
Random effects:
 Groups  Name        Var ance Std.Dev.
 country (Intercept) 0.2543   0.5042
Number of obs: 27563, groups:  country, 30
```

Group variance

$$ICC = \frac{0.2543}{0.2543 + \frac{\pi^2}{3}} = 0.072$$

# Fit the random intercept model

```
> library(sjPlot)
> m1 = glmer(cultinc ~ eduscaled + agescaled + gender + bill + (1|country),
          data = cult1, family = binomial)
> tab_model(m1)
```

| | cultinc | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 0.96 | 0.79 – 1.18 | 0.707 |
| eduscaled | 2.42 | 2.32 – 2.52 | **<0.001** |
| agescaled | 0.40 | 0.39 – 0.42 | **<0.001** |
| gender [Female] | 0.99 | 0.93 – 1.05 | 0.772 |
| bill [no] | 1.48 | 1.38 – 1.58 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ country}$ | 0.28 | | |
| ICC | 0.08 | | |
| $N_{country}$ | 30 | | |
| Observations | 24499 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.330 / 0.384 | | |

# Fit the random intercept and slope model

Let's randomize the coef for age:

```
> m2 = glmer(cultinc ~ eduscaled + agescaled + gender + bill + (1+agescaled|country),
             data = cult1, family = binomial)
> tab_model(m2)
```

| | cultinc | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 0.94 | 0.78 – 1.15 | 0.569 |
| eduscaled | 2.46 | 2.36 – 2.56 | **<0.001** |
| agescaled | 0.40 | 0.37 – 0.44 | **<0.001** |
| gender [Female] | 1.01 | 0.95 – 1.07 | 0.855 |
| bill [no] | 1.45 | 1.36 – 1.55 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ country | 0.27 | | |
| $\tau_{11}$ country.agescaled | 0.05 | | |
| $\rho_{01}$ country | 0.39 | | |
| ICC | 0.09 | | |
| N country | 30 | | |
| Observations | 24499 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.333 / 0.393 | | |

# Let’s compare

```
> anova(m2, m1)

Data: cult1
Models:
m1: cultinc ~ eduscaled + agescaled + gender + bill + (1 | country)
m2: cultinc ~ eduscaled + agescaled + gender + bill + (1 + agescaled | country)
   Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
m1  6 26270 26318 -13129    26258
m2  8 26157 26222 -13071    26141 116.41      2  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The second model is significantly better

# Add a country-level variable

Let's randomize the coef for age:

```
> library(optimx)
> m3 = glmer(cultinc~ eduscaled + agescaled + gender + bill + loggdp +
(1+agescaled|country), data = cult1, family = binomial,
glmerControl(optimizer = "optimx", calc.derivs = FALSE,
optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
```

# Add a country-level variable

Let's randomize the coef for age:

```
> library(optimx)
> m3 = glmer(cultinc~ eduscaled + agescaled + gender + bill + loggdp +
(1+agescaled|country), data = cult1, family = binomial,
glmerControl(optimizer = "optimx", calc.derivs = FALSE,
optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
```

This helps your model to converge (use if the model fails to converge)

# Add a country-level variable

```
> tab_model(m2)
```

| | cultinc | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 0.00 | 0.00 – 0.00 | **<0.001** |
| eduscaled | 2.46 | 2.36 – 2.56 | **<0.001** |
| agescaled | 0.40 | 0.36 – 0.44 | **<0.001** |
| gender [Female] | 1.02 | 0.96 – 1.08 | 0.493 |
| bill [no] | 1.45 | 1.36 – 1.56 | **<0.001** |
| loggdp | 2.16 | 1.75 – 2.65 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ country}$ | 0.11 | | |
| $\tau_{11\ country.agescaled}$ | 0.05 | | |
| $\rho_{01\ country}$ | -0.27 | | |
| ICC | 0.04 | | |
| $N_{country}$ | 29 | | |
| Observations | 24048 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.378 / 0.405 | | |

# Add a cross-level interaction variable

```
> m4 = glmer(cultinc~ eduscaled + agescaled + gender + bill + loggdp + loggdp:agescaled +
(1+agescaled|country), data = cult1, family = binomial,
glmerControl(optimizer = "optimx", calc.derivs = FALSE,
optCtrl = list(method = "nlminb", starttests = FALSE, kkt = FALSE)))
tab_model(m4)

> tab_model(m4)
```

# Add a cross-level interaction variable

| | cultinc | | |
|---|---|---|---|
| *Predictors* | *Odds Ratios* | *CI* | *p* |
| (Intercept) | 0.00 | 0.00 – 0.01 | **<0.001** |
| eduscaled | 2.46 | 2.36 – 2.56 | **<0.001** |
| agescaled | 0.04 | 0.01 – 0.14 | **<0.001** |
| gender [Female] | 1.02 | 0.96 – 1.09 | 0.490 |
| bill [no] | 1.45 | 1.36 – 1.55 | **<0.001** |
| loggdp | 1.99 | 1.61 – 2.45 | **<0.001** |
| agescaled * loggdp | 1.27 | 1.11 – 1.45 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ country}$ | 0.10 | | |
| $\tau_{11\ country.agescaled}$ | 0.03 | | |
| $\rho_{01\ country}$ | -0.24 | | |
| ICC | 0.04 | | |
| $N_{country}$ | 29 | | |
| Observations | 24048 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.379 / 0.402 | | |

# Interpretation

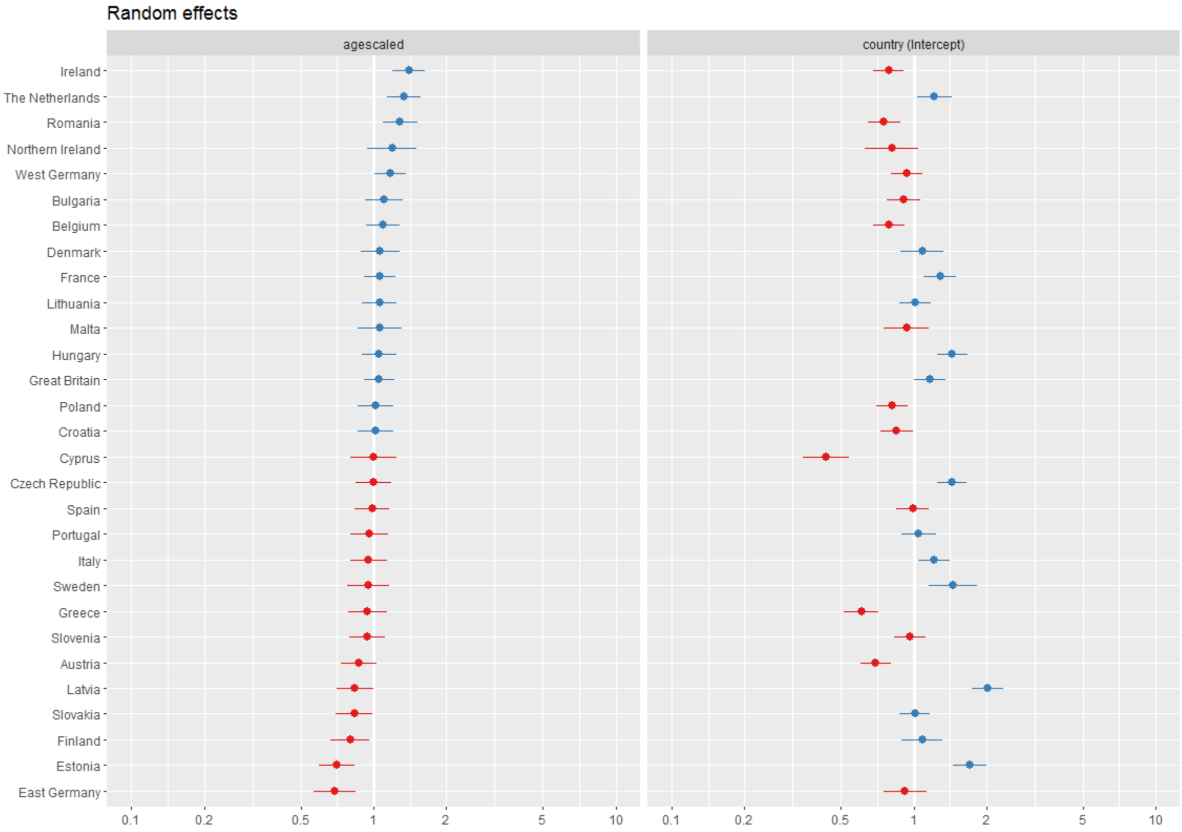
- All variables except gender are significantly related to the probability of internet use for cultural purposes.
- Age when the respondent completed education is positively related to the probability of internet use for cultural purposes. 1 standard deviation increase doubles the probability of internet use for cultural purposes.
- Age is negatively related to the probability of internet use for cultural purposes. 1 standard deviation increase in age leads to 96% decrease in the probability of internet use for cultural purposes.
- People who don’t have difficulties paying bills have 45% more likely to use internet for cultural purposes.
- GDP per capita is positively related to the probability of internet use for cultural purposes. 1-unit increase in the log od GDP doubles the probability internet use for cultural purposes.
- The effect of age depends on the wealth of the country. 1-unit increase in the log od GDP increases the effect of age by 1.27. As the effect of age is negative, it means that the growth of GDP leads to the decline in age differences.
- The model fit is excellent (Look at R2)

# Let’s plot

```
> plot_model(m4, type = 're',
sort.est = 'agescaled')
```

Ireland has the smallest age differences

East Germany has the biggest age gap



# Let’s plot

```
> plot_model(m4, type = ‘int')
```

You can see that as GDP growth the age gap becomes smaller