STATISTICAL HORIZONS

Multilevel and Mixed Models Using R

Stephen Vaisey, Ph.D.

Upcoming Seminar:

July 28-31, 2020, Remote

By the end of tomorrow, you will…

- Get the "big picture" intuition of MLMs
- Understand what random intercepts and random slopes are when to use each one
- Know the difference between fixed and random effects and how to implement models that use both
- Know how to estimate MLMs (linear and logit) in R and interpret their results

7

7

What are MLMs?

Most broadly, they are models that estimate parameters from data with:

- multiple observations from the same group (e.g., students in classes)
- repeated observations from the same person
  - sometimes called panel data or longitudinal data
  - not the primary focus here but *very similar*

8

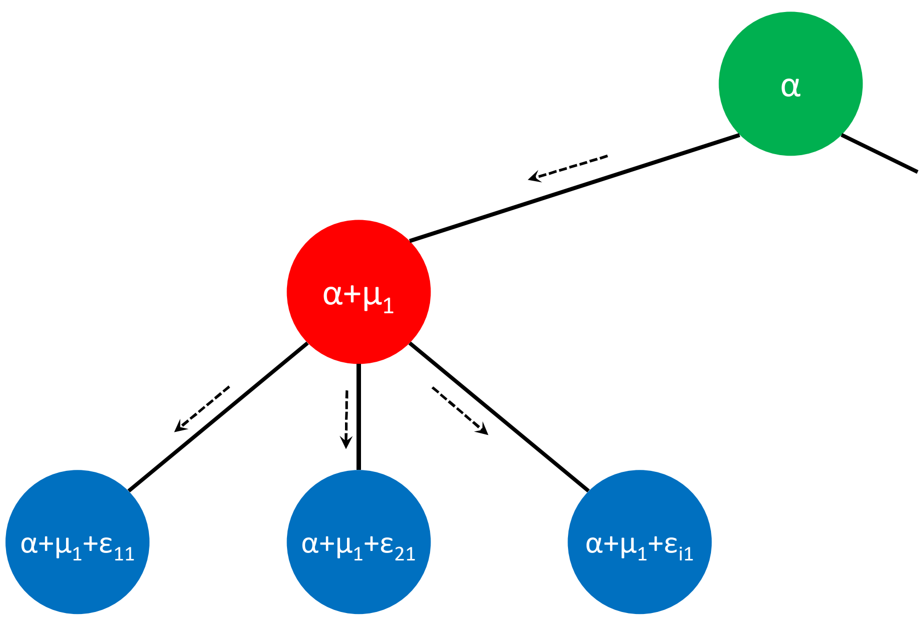
8

Adding α for the grand mean

$$y_{ij} = \alpha + \mu_j + \varepsilon_{ij}$$

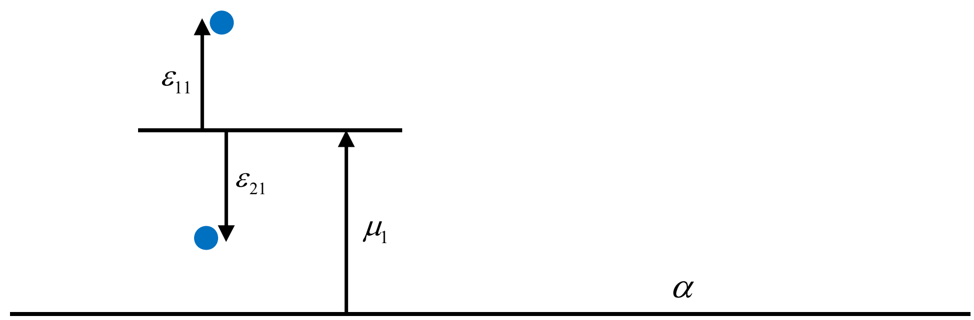
19

19



20

20



21

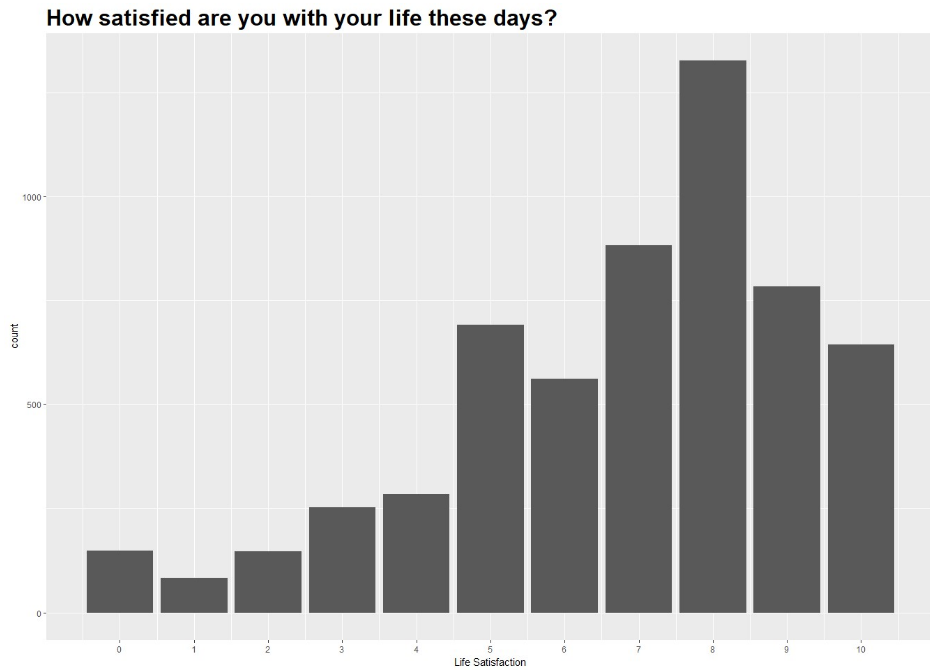
21

Sometimes written in two equations
with distributional assumptions spelled out

$$\text{Level 1}: y_{ij} = \beta_{0j} + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2)$$
$$\text{Level 2}: \beta_{0j} = \beta_0 + \mu_j, \mu_j \sim N(0, \tau^2)$$

22

22



27

27

```
> describe(ess_simple$stflife)
   vars    n mean   sd median trimmed  mad min max range  skew kurtosis   se
X1    1 5800 6.78 2.41      7    7.02 2.97   0  10    10 -0.86      0.3 0.03
> table(ess_simple$cntry)

 AL  BE  BG  CH  CY  CZ  DE  DK  EE  ES  FI  FR  GB  HU  IE  IL  IS  IT  LT  NL  NO  PL  PT  RU  SE  SI  SK  UA  XK 
200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 
```

Life Satisfaction

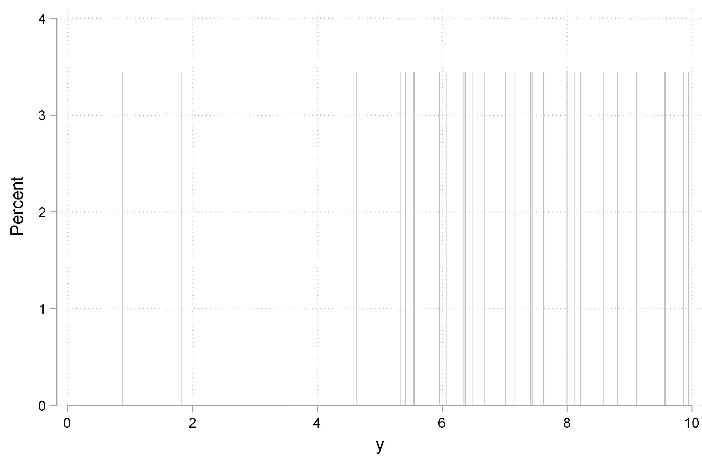
Mean	6.8
Variance	5.8

(remember these numbers)

28

28

Two paths to the same variance: (1) all at the country level



29

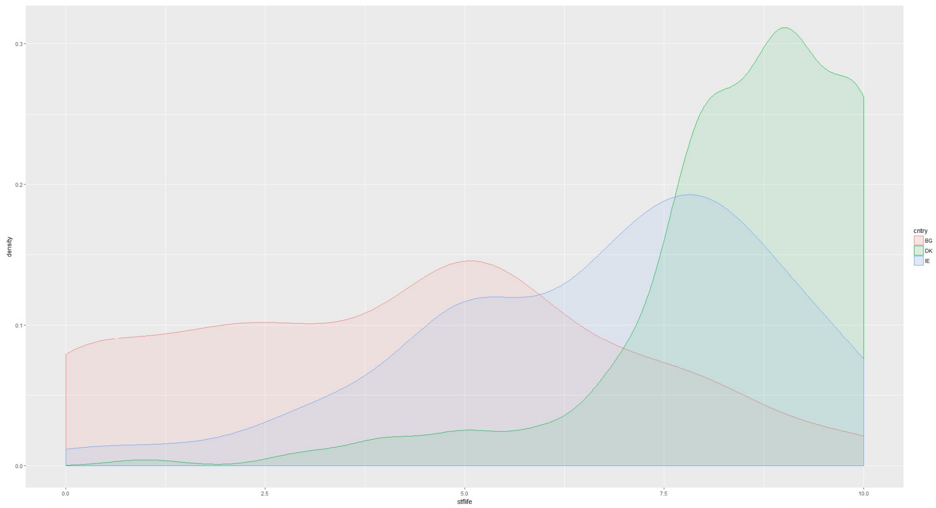
29

Two paths to the same variance: (2) all at the individual level

30

30

(As usual) a mix of both



31

31

Intraclass Correlation (ρ)

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2}$$

$$ICC = \frac{\text{L2 variance}}{\text{total variance}}$$

$$\rho = \frac{\psi}{\psi + \theta}$$

32

32

The simplest mixed model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + \varepsilon_{ij}$$

$$sat_{ij} = \beta_0 + \beta_1 ed_{ij} + u_j + \varepsilon_{ij}$$

63

Basic `lme4` syntax

```
lmer(y ~ x + (1 | cluster), # cluster is L2 id
     data = df,             # df is data frame
     REML = FALSE)          # use maximum likelihood
```

64

```
> # Basic mixed model
> m2 <- lmer(stflife ~ eduyrs + (1 | cntry),
+            data = ess_simple,
+            REML = FALSE)
> summary(m2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stflife ~ eduyrs + (1 | cntry)
   Data: ess_simple

     AIC      BIC   logLik deviance df.resid
 25661.9  25688.6 -12827.0  25653.9     5796

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.7077 -0.5339  0.1210  0.6780  2.6587

Random effects:
 Groups   Name        Variance Std.Dev.
 cntry    (Intercept) 0.9607   0.9801
 Residual             4.7905   2.1887
Number of obs: 5800, groups:  cntry, 29

Fixed effects:
            Estimate Std. Error t value
(Intercept)  6.20161    0.20752   29.88
eduyrs       0.04716    0.00773    6.10
```


	Null model	Null + education	
Fixed part			
Years of education		0.047	
Constant	6.784	6.202	
Random part			
L2 variance	1.009	0.961	
L1 variance	4.820	4.791	
ICC	.173	.167	This is now a *residual* ICC which isn't very useful

Adding random slopes: What *are* they?

Random slopes (a.k.a. random coefficients) are interactions between **observed** L1 variables and **unobserved** L2 variables.

They allow L1 slopes to differ between L2 clusters without accounting for **why** these differences might exist.

77

77

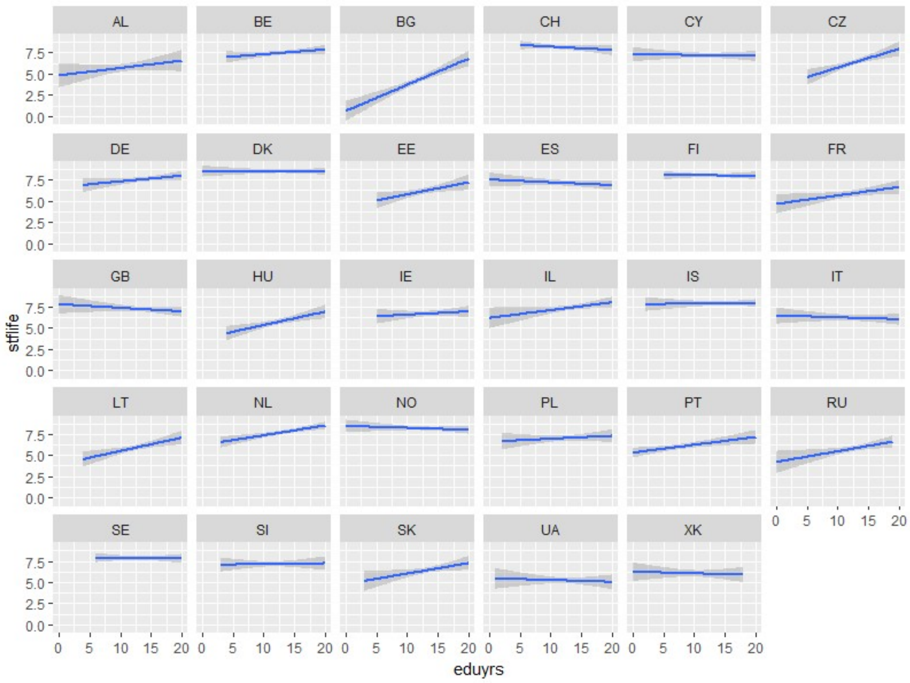
Random coefficients/slopes

$$sat_{ij} = \underbrace{b_0 + u_{0j}}_{\text{intercept}} + (\underbrace{b_1 + u_{1j}}_{\text{slope}})ed_{ij} + \varepsilon_{ij}$$

A cluster-specific increment (or decrement) is added to (subtracted from) the "grand mean" (b_1) of the slope

78

78



Looking at a graph like this might make us wonder if the effects differ systematically by country!

79

79

Random coefficients/slopes syntax

```
lmer(y ~ x + (1 + x | cluster),   # x is now in both EQs
     data = df,
     REML = FALSE)
```

80

80

The Between-Within (BW) method

For one within-cluster varying variable, *x*, (e.g., years of education) and one cluster-invariant variable, *z*, (e.g., GNI per capita), it would look like this:

$$y_{ij} = \beta_0 + \underbrace{\beta_1(x_{ij} - \overline{x}_j)}_{\text{within}} + \underbrace{\beta_2\overline{x}_j + \beta_3 z_j}_{\text{between}} + u_j + \varepsilon_{ij}$$

123

123

BW model results

```
> summary(bw1)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stflife ~ b.eduyrs + w.eduyrs + (1 | cntry)
   Data: ess_simple_bw

     AIC      BIC   logLik deviance df.resid
   25659    25692   -12824    25649     5795

Scaled residuals:
   Min      1Q Median     3Q    Max
-3.702 -0.530  0.118  0.677  2.657

Random effects:
 Groups   Name        Variance Std.Dev.
 cntry    (Intercept) 0.799    0.894
 Residual             4.790    2.189
Number of obs: 5800, groups:  cntry, 29

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.06758    1.74312    1.19
b.eduyrs     0.38192    0.14049    2.72
w.eduyrs     0.04631    0.00774    5.98
```

the w.eduyrs coefficient is *exactly* the same as any FE estimate

124

124

	How satisfied with life as a whole	How satisfied with life as a whole
	B (CI)	*B (CI)*
Fixed Parts		
(Intercept)	6.784 (6.41 – 7.15) ***	2.068 (-1.35 – 5.48)
Education (between)		0.382 (0.11 – 0.66) **
Education (within)		0.046 (0.03 – 0.06) ***
Random Parts		
σ^2	4.820	4.790
$\tau_{00,\ cntry}$	1.009	0.799
N_{cntry}	29	29
ICC_{cntry}	0.173	0.143
Observations	5800	5800
Notes		* *p*<.05 ** *p*<.01 *** *p*<.001

$$R^2_{L1} = \frac{4.820 - 4.790}{4.820} = .006$$

$$R^2_{L2} = \frac{1.009 - .799}{1.009} = .208$$

$$R^2_{overall} = \frac{(4.820 + 1.009) - (4.790 + .799)}{(4.820 + 1.009)} = .041$$

Mean education accounts for 21% of the between-country variance in life satisfaction. Within-country education differences don't account for much within-country variance.

125

125

Adding L2-only variables

```
> bw2 <- lmer(stflife ~ b.eduyrs + w.eduyrs + gnipc1000 + (1 | cntry),
+             REML = FALSE,
+             data = ess_simple_bw)
> summary(bw2)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: stflife ~ b.eduyrs + w.eduyrs + gnipc1000 + (1 | cntry)
   Data: ess_simple_bw

     AIC      BIC   logLik deviance df.resid
   25638    25678   -12813    25626     5794

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.726 -0.525  0.120  0.674  2.639

Random effects:
 Groups   Name        Variance Std.Dev.
 cntry    (Intercept) 0.357    0.597
 Residual             4.790    2.189
Number of obs: 5800, groups:  cntry, 29

Fixed effects:
            Estimate Std. Error t value
(Intercept) 3.58579   1.21396    2.95
b.eduyrs    0.18010   0.10167    1.77
w.eduyrs    0.04631   0.00774    5.98
gnipc1000   0.03111   0.00536    5.81
```

You don't have to do anything special to include L2 variables. Just add them into the fixed portion like any other variable.

126

126

Comparing many models

Many decisions to make, for example:

- Do I need to split out education into between/within?
- Do I need to include a specific variable at all?
- Do I need random slopes on the L1 variables?

129

129

AIC/BIC of all models (so far)

Model	Specification	AIC	BIC
M1	random intercepts	25,697	25,717
M2	RI + eduyrs	25,662	25,689
M3	RI + eduyrs*	25,610	25,650
BW1	RI + b/w.eduyrs	25,659	25,692
BW2	RI + b/w.eduyrs + GDP	25,638	25,678
BW3	RI + b/w*.eduyrs + GDP	**25,592**	25,645
M4	RI + eduyrs* + GDP	**25,592**	**25,638**

* signifies random slopes

130

130

Cross-level interactions

We might want to know whether the *context* of a cluster (i.e., cluster-level attributes) conditions the effects of a within-cluster difference.

For example, do within-country educational differences matter *less* for determining life satisfaction in richer countries?

133

133

Centering and interactions

We center variables to make their related coefficients more interpretable.

If done properly, centering has no effect on **model fit** but facilitates **model interpretation**.

134

134

Limited Dependent Variables

149

149

What can we do with `lme4`?

```
binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

150

150