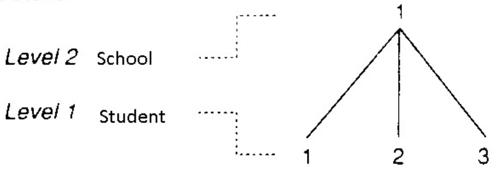
Introduction to multi-level models

Instructor: Gabriel Katz

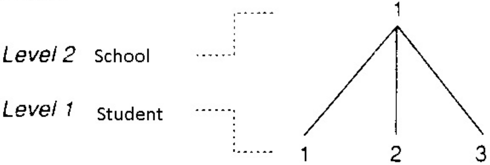
Introduction to Multilevel Models

- Assume you are interested in studying the determinants of students' test scores
- It may be reasonable to assume that certain characteristics of the students influence their score
  - ethnicity, parent's background and socio-economic status, effort (number of hours devoted to studying), etc.
- In addition, one might think that school characteristics also affect students' scores
  - neighbourhood, whether student/teacher ratio, whether public or private, etc.
- So, one could argue that we have in fact two level of analysis:
  1. individual students
  2. schools the students attend

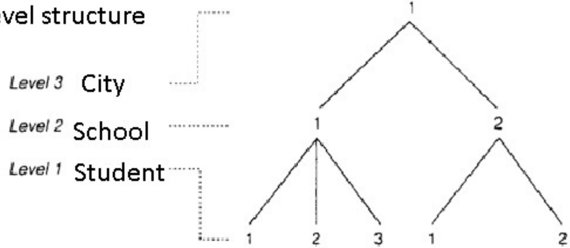
Two-level structure



Two-level structure



Three-level structure



Other examples of multi-level or hierarchical data structures

- Politics: Voters within districts
    - vote choice influenced by individual characteristics: partisanship, age, education, etc.
    - but also by contextual factors: city, neighbourhood, social networks
- Criminology: racial profiling
    - individual may be stopped by the police because of their personal characteristics (age, "looks", etc.)
    - but also broader factors (e.g., ethnic minority groups)
- Panel/longitudinal data?
- Cross-nested models: random effects for different clustering units

Characteristics of multilevel data

- Clustered or nested structure
  - e.g., students within schools, voters within districts, individuals within "race"
- Heterogeneity and dependence
  - the behaviour/attitudes/opinions of individual within the same cluster/group may be correlated
  - behaviours/attitudes/opinions may vary across clusters/groups
- The researcher may want to take into account the impact that factors "at different levels" have on behaviours/opinions/attitudes
  - i.e., account for micro & macro relations

Multilevel or hierarchical models are good for these problems

- Let's go back to the example of students' scores ... and use some math (just a little bit)
- I am interested in studying the determinants on student i's score, y_i , $i = 1, \ldots, n$
- Specifically, I am interested in:
  - A certain characteristic of i - e.g., time spent studying - denoted as x_i
  - A certain characteristic of the school, j, where i studies
    - e.g., the student/teach ratio in school j, denoted as z_j

- Assume - for now - that I think that the school influences i's average score.
- Then, a multi-level model for this problem could be written as:

$$\text{Level 1:} \qquad y_i = \alpha_j + \beta x_i + \epsilon_i \tag{1}$$

- (1) looks like a standard regression: the only difference is the intercept:
  - we would usually write it as α
  - here, however, we write it as α_j
- So we are allowing the intercept - the average student score - to vary across schools
  - e.g., students in school A get better grades *on average* than students in school B

- So the influence of schools on students' marks "operates" through α_j
- Moreover, I assumed that the impact that schools have on students' grades depends on a specific characteristic: teacher/student ratio, z_j
- That is, the impact of school on students is in turn affected by z_j:

- So the influence of schools on students' marks "operates" through α_j
- Moreover, I assumed that the impact that schools have on students' grades depends on a specific characteristic: teacher/student ratio, z_j
- That is, the impact of school on students is in turn affected by z_j:

$$\text{Level 2:} \qquad \alpha_j = \gamma + \delta z_j + \nu_j \tag{2}$$

- This is a "new" (second-level) regression
- Together, (1) and (2) form a multilevel model

$$\begin{aligned} y_i &= \alpha_j + \beta x_i + \epsilon_i && \text{for student} \quad i = 1, \ldots, n \\ \alpha_j &= \gamma + \delta z_j + \nu_j && \text{for school} \quad j = 1, \ldots, J \end{aligned}$$

Extending the example

- So far, we assume that schools only affect the average marks of its students
- The researcher might believe, however, that schools mediate the impact that a student's individual effort has on her mark
  - i.e., in a school with low student/teacher ratio, students are more "effective"
  - thus, the time spent by students preparing for the test is "worth more" than in schools with high student/teacher ratio
- In other words, the influence of schools on students' marks "operates" not through the intercept α, but through the coefficient of x_i

- The new multilevel model could be written as:

$$\begin{aligned} y_i &= \alpha + \beta_j x_i + \epsilon_i \qquad \text{for student} \quad i = 1, \ldots, n \\ \beta_j &= \delta + \theta z_j + \eta_j \qquad \text{for school} \quad j = 1, \ldots, J \end{aligned}$$

- Alternatively, one might believe that the influence of the school operates *via* 2 ways:

  1. affecting students' average marks
  2. affecting the effectiveness/efficiency of students' effort

- The new multilevel model could be written as:

$$\begin{aligned} y_i &= \alpha + \beta_j x_i + \epsilon_i && \text{for student} \quad i = 1, \ldots, n \\ \beta_j &= \delta + \theta z_j + \eta_j && \text{for school} \quad j = 1, \ldots, J \end{aligned}$$

- Alternatively, one might believe that the influence of the school operates *via* 2 ways:

  1. affecting students' average marks
  2. affecting the effectiveness/efficiency of students' effort

$$\begin{aligned} y_i &= \alpha_j + \beta_j x_i + \epsilon_i && \text{for student} \quad i = 1, \ldots, n \\ \alpha_j &= \gamma + \delta z_j + \nu_j && \text{for school} \quad j = 1, \ldots, J \\ \beta_j &= \lambda + \theta z_j + \eta_j && \text{for school} \quad j = 1, \ldots, J \end{aligned}$$

- Note that in multilevel models, we have several "error terms" reflecting factors not captured by the variables x_i and z_j
- ϵ_i is the "standard" error term in conventional regression models: reflects individual-level factors influencing y_i beyond x_i
  - ϵ_i captures unobserved factors influencing y_i at the individual level
- η_j and ν_j are error terms at the school level
  - captures other factors influencing the impact that schools have on students' marks beyond z_j
  - without η_j and ν_j, we would be assuming that all the influence that schools have on y_i is perfectly measured by z_j
  - also note that two students attending the same school j will have different ϵs, but the same ν_j and η_j (and, in fact, the same α_j and β_j)
  - $\rightarrow$ correlation between students in the same school

- Typically, we attribute normal distributions to the error terms:

$$\epsilon \sim_{iid} N(0, \sigma^2_\epsilon)$$
$$\eta \sim_{iid} N(0, \sigma^2_\eta)$$
$$\nu \sim_{iid} N(0, \sigma^2_\nu)$$

- The variances σ^2_ϵ, σ^2_η, σ^2_ν measure the variability or dispersion in the data
  - σ^2_ϵ measure the (unexplained, unmeasured) variability at Level 1 (as in usual regression); within Level 2 units
  - σ^2_ν measures the unexplained variation in α_j, between Level 2 units
  - σ^2_η measures the unexplained variation in β_j, between Level 2 units
- So, for instance, the larger σ^2_ν, the larger the variability across Level 2 units
- Other assumptions besides normality - e.g., Student-t

Alternative approaches to this problem

- Before you knew about multi-level models, how did you address this research problem?

Alternative approaches to this problem

- Before you knew about multi-level models, how did you address this research problem?

  1. Separate regressions for each school:
     - run a "single-level" regression for each school; examine variation across schools to detect "school effect"
     - **problem**: if some schools have few students, the estimates are going to be very imprecise; not a lot of information
     - might not be a big problem for schools; but how about districts, cities, ethnic groups in a survey, etc.
  2. Include school level predictors in the (single-level) individual regression
     - **problem**: we assume that the school's influence is perfectly measured by z_j
     - two different schools with the same value of z_j will have the same impact on y_i

3. Add both school-level covariates and school dummies to account for unmeasured school influence
   - **problem**: (multi) collinearity between school dummies and school variables
4. Two-stage regression:
   - first include school dummies in the individual regression
   - second stage: regress dummy coefficient on school covariates
   - **problem 1**: if few students per school, estimates of dummy coefficients won't be very good (same as in alternative 1)
   - **problem 2**: ignoring error terms of the dummy estimates in stage 1 - "spurious significance"

- Multilevel models avoid all these problems
  - in particular, very useful when few level-1 units (students) within level-2 units (schools)
  - much simpler than two-stage approaches

Other names for multilevel models

- Random-effects models
- Hierarchical models
- Variance-components models
- Mixed models

Fitting multilevel models

- Multilevel models can be estimated using different statistical software packages
  - Stata
  - SAS
  - MLWin
  - HLM
  - R
  - WinBUGS (JAGS)
- Today, we are going to use R. Why R?
  1. It's free!
  2. Lots of packages readily available covering a wide array of "topics"
  3. Becoming increasingly used in the social sciences
  4. Works "alone", but also easy to use with WinBUGS
- Find all about R: http://www.r-project.org/

Multilevel models in R

- Two basic commands to fit multilevel models in R
  - *lmer*, for multilevel linear regression
  - *glmer*, for generalized linear multilevel regression (e.g., multilevel logit, probit)
- We are going to analyze data from my own research:
  - "Peru.csv": data on presidential approval in Peru
  - 2 levels: individuals (rows), and province
- We are going to fit a series of multilevel models
- Let's take a minute to take a look at the data

The first 4 columns are "Level-1" (individual survey respondents).
The other 4 are "Level-2" (province)

president.approval	age	education	indigenous	pop.density	illiteracy	infant.mo	province
NA	67	2	1	148.127471	0.0928	0.128	21801
0	36	4	0	148.127471	0.0928	0.128	21801
0	26	5	0	148.127471	0.0928	0.128	21801
0	34	4	0	148.127471	0.0928	0.128	21801
0	25	4	0	148.127471	0.0928	0.128	21801
0	54	3	1	148.127471	0.0928	0.128	21801
0	30	4	0	148.127471	0.0928	0.128	21801
0	33	8	0	148.127471	0.0928	0.128	21801
0	39	9	0	148.127471	0.0928	0.128	21801
1	40	9	0	148.127471	0.0928	0.128	21801
1	67	6	1	148.127471	0.0928	0.128	21801
0	58	3	1	148.127471	0.0928	0.128	21801
0	40	7	NA	148.127471	0.0928	0.128	21801
0	60	5	0	148.127471	0.0928	0.128	21801
0	29	8	0	148.127471	0.0928	0.128	21801
0	44	3	NA	148.127471	0.0928	0.128	21801
0	21	8	0	148.127471	0.0928	0.128	21801
0	31	4	0	148.127471	0.0928	0.128	21801
0	75	4	1	148.127471	0.0928	0.128	21801
0	34	9	0	148.127471	0.0928	0.128	21801

- The dependent variable is dichotomous: $1 = \textit{approve}, 0 = \textit{disapprove}$
- Forget for a minute that dep. variable is binary
  - treat it as continuous (for the moment)
  - $\rightarrow$ multilevel linear model $\rightarrow$ command:*lmer*
- We are going to start with the most basic multilevel model, M.1
  - regression with no individual level covariates and province-level intercepts

$$\begin{array}{ll} \text{Level 1} & y_i = \alpha_j + \epsilon_i \\ \text{Level 2} & \alpha_j = \delta + \nu_j \end{array}$$

- This multilevel model can be written as:

$$y_i = \delta + \nu_j + \epsilon_i$$

(just replace α_j in Level 1)

- How do we write this in R using *lmer*?

- How do we write this in R using *lmer*?

```
M.1<-lmer(president.approval~1+(1|province))
```

- Compare again with:

$$y_i = \delta + \nu_j + \epsilon_i$$

- Here:
  - `~ is the R symbol for ''equal to''`
  - `1 is a constant term corresponding to` δ
  - `(1|province)` tells R that we want an error term at level 2, that varies by province

How to interpret the results

- We can take a look at the results:

```
display(M.1)
fixef(M.1)
ranef(M.1)
coef(M.1)
```

- Two type of coefficients in our model:
  - δ: average presidential approval; fixed effects: common to all units in the sample; no variation across groups
  - $\alpha_j = \delta + \nu_j$: specific to each province: "coef"
  - $\nu_j = \alpha_j - \delta$ measures how much province j deviates from the average approval rate: "ranef" f"
    - if $\nu_j > 0 \rightarrow \alpha_j > \delta$
    - if $\nu_j < 0 \rightarrow \alpha_j < \delta$

- Note that, in the first province in our data, there are province-level factors that drive down the average support for the president:

```
 > fixef(M.1)
(Intercept)
  0.2807867

> ranef(M.1)$province[1,1]
[1] -0.09469842

> fixef(M.1)+ranef(M.1)$province[1,1]
(Intercept)
  0.1860883

> coef(M.1)$province[1,1]
[1] 0.1860883
```

- Since $\nu_1 < 0$, $\alpha_1 = \delta + \nu_1 < \delta$

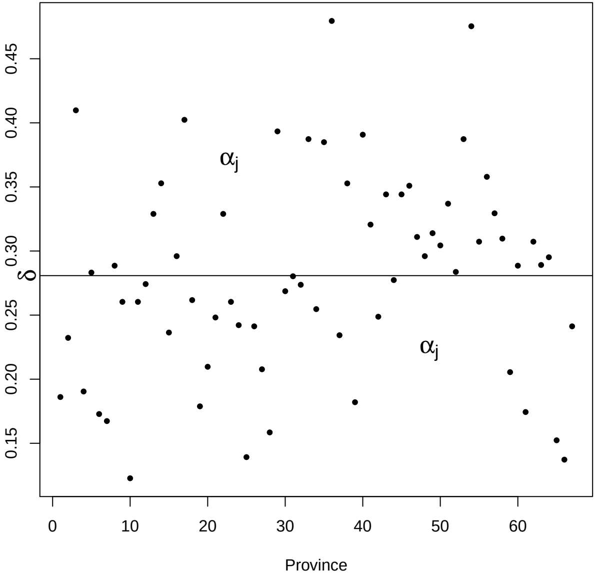
- In contrast, some factors in province 3 increase the average support for the president

```
> ranef(M.1)$province[3,1]
[1] 0.1289884

> fixef(M.1)+ranef(M.1)$province[3,1]
(Intercept)
  0.4097751

> coef(M.1)$province[3,1]
[1] 0.4097751
```

- So, $\alpha_3 > \delta$



0.45
0.40
0.35
0.30
δ
0.25
0.20
0.15
αj
αj
0
10
20
30
40
50
60
Province

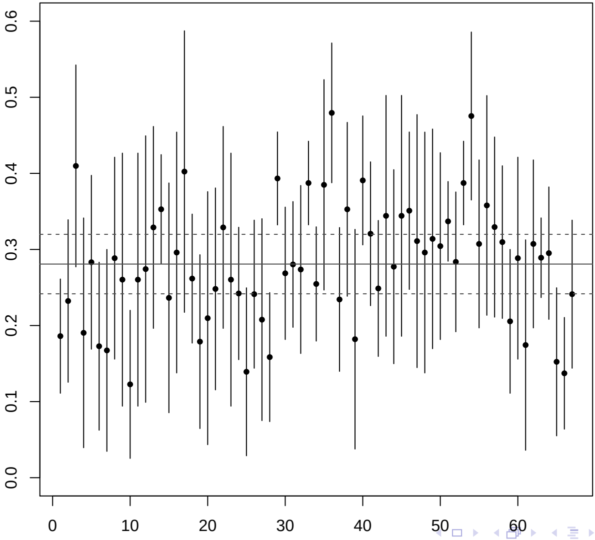
It is a little bit more complicated!

- So far, we have only worked with point estimates; $\hat{\delta}$, $\hat{\alpha}_j$
- But these quantities are random variables: they have some associated uncertainty
  - just as in the "standard" regression, we have standard errors for δ and α_j
- In R, we can access the standard error of the fixed and random effects:

```
> se.fixef(M.1)
(Intercept)
 0.02004559

> se.ranef(M.1)
```

Now the figure with standard errors



- We can also take a look at the standard errors:

```
> display(M.1)
...
Error terms:
 Groups   Name        Std.Dev.
 province (Intercept) 0.12
 Residual             0.44
```

- This gives us the standard errors at Levels 1 and 2: $\hat{\sigma}_\epsilon$ and $\hat{\sigma}_\nu$

- Note that the total unexplained variation in presidential approval is given by:

$$\hat{\sigma}^2_\epsilon + \hat{\sigma}^2_\nu = 0.44^2 + 0.12^2 = 0.208$$

- Level-2 variation explains 7% of the total variation in presidential approval:

$$\frac{\hat{\sigma}^2_\nu =}{\hat{\sigma}^2_\epsilon + \hat{\sigma}^2_\nu} = \frac{0.12^2}{0.208} = 0.07$$

Let's add a level-1 variable

- We include individual's ethnicity in the regression, x_i: a dummy for "indigenous"
- We now have Model.2:

$$y_i = \alpha_j + \beta x_i + \epsilon_i$$
$$\alpha_j = \delta + \nu_j$$

- or, in just 1 line:

$$y_i = \delta + \nu_j + \beta x_i + \epsilon_i$$

- Let's write this model in R:

```
M.2<-lmer(president.approval~indigenous+(1|province))
```

- Note that we don't write

```
M.2<-lmer(president.approval~1+indigenous+(1|province))
```

- Let's interpret the results:
  - β and δ are the same for all units: these are fixed effects

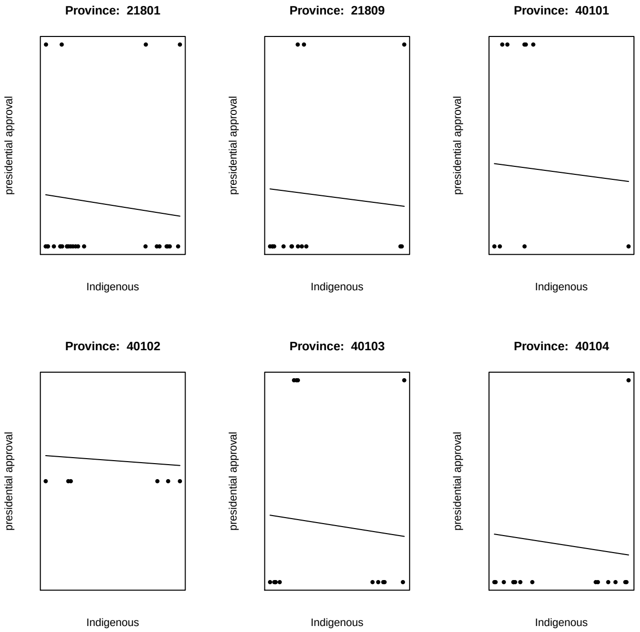
```
> fixef(M.2)
(Intercept)  indigenous
 0.31379260 -0.09926413
```

  - ν_j are error terms that vary across provinces

```
> ranef(M.2)$province
        (Intercept)
21801  -0.073574089
21809  -0.042213587
40101   0.083094120
40102  -0.065010353
40103   0.006253472
...
```

  - thus, α_j also varies across provinces

```
> coef(M.2)$province
       (Intercept)  indigenous
21801    0.2402185 -0.09926413
21809    0.2715790 -0.09926413
40101    0.3968867 -0.09926413
```


Province: 21801
presidential approval
Indigenous
Province: 21809
presidential approval
Indigenous
Province: 40101
presidential approval
Indigenous
Province: 40102
presidential approval
Indigenous
Province: 40103
presidential approval
Indigenous
Province: 40104
presidential approval
Indigenous

Let's include level-2 variables now!

- We now add a province-level predictor: illiteracy rate, z_j.
- The model is:

$$y_i = \alpha_j + \beta x_i + \epsilon_i$$
$$\alpha_j = \delta + \gamma z_j + \nu_j$$

- or, in just 1 line:

$$y_i = \delta + \gamma z_j + \beta x_i + \epsilon_i + \nu_j$$

- In R:

```
M.3<-lmer(president.approval~illiteracy+
indigenous+(1|province))
```

- We can see the estimated "fixed effects", $\hat{\delta}$, $\hat{\gamma}$, $\hat{\beta}$

```
> fixef(M.3)
(Intercept)  indigenous  illiteracy
 0.58429854 -0.07864703 -3.40614062
```

- As well as the province-level error terms

```
> ranef(M.3)$province
         (Intercept)
21801  -0.0325753412
21809  -0.0244399499
40101   0.0204332684
40102  -0.0353049912
40103   0.0034263206
40104  -0.0449360147
...
```

- Let's interpret these results ...

- It is also interesting to see what happened with the Level-1 and Level-2 error terms when we added province-level predictors.
- Remember that in Model 2 we only had Level-1 variables and Level-2 intercepts, but no Level-2 covariate.

```
> display(M.2)
...
Error terms:
 Groups   Name        Std.Dev.
province (Intercept) 0.10
 Residual             0.44
---

> display(M.3)
..
Error terms:
 Groups   Name        Std.Dev.
 province (Intercept) 0.07
 Residual             0.44
---
```

- Note that the Level-1 standard error did not change: $\hat{\sigma}_\epsilon = 0.44$
- It makes sense, since we did not add any new Level-1 variable
- However, the Level-2 standard error became smaller
  - For M.2, $\hat{\sigma_\nu} = 0.10$
  - For M.3, $\hat{\sigma_\nu} = 0.07$
- Remember that the error terms capture everything not measured by the covariates
  - and $\hat{\sigma_\nu}$ captures the dispersion in these unmeasured effects across Level-2 units
- In this case, ν_j captures unobserved factors that are influencing α_j
  - or, ultimately, that province-level factors that are influencing y_i through α_j
- When we added z_j, the "unmeasured" province effect on y_i became smaller

Now let's allow β to vary across provinces

- So far, we have assumed that province-level factors only influence the average presidential job approval

- However, province-level factors might also influence β, i.e., the impact of being a native american on job approval

- To do so, we need to express β as a random term varying across provinces, $j = 1, \ldots, J$.

- Let's start with M.4, a model with no province-level covariates:

$$y_i = \alpha_j + \beta_j x_i + \epsilon_i$$
$$\alpha_j = \delta + \nu_j$$
$$\beta_j = \lambda + \eta_j$$

- We can also write M.4 as:

$$y_i = \delta + \lambda x_i + \eta_j x_i + \nu_j + \epsilon_i$$

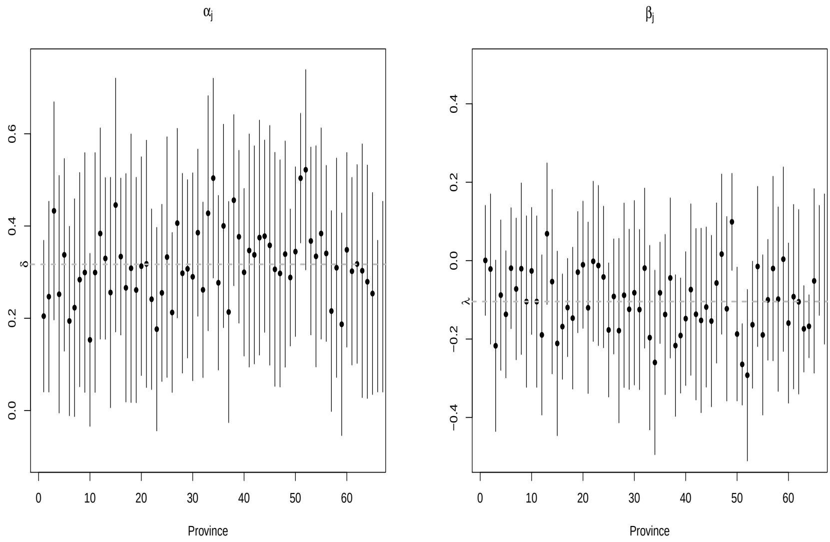
- What the is formula for this in R?

```
M.4<-lmer(president.approval~indigenous+
 (1+indigenous|province))
```

- So in this model we have:
  - two fixed effects that do not vary across groups: the constant associated with δ and λ
  - two Level-2 random terms, one corresponding to α_j and the other corresponding to β_j, which multiplies x_i (indigenous)
  - the two random terms are captured in this part of the formula: `(1+indigenous|province))`

- Note that now, when we ask R for the model's coefficients, both the intercept and the coefficient of *indigenous* vary across provinces:

```
 coef(M.4)
$province
       (Intercept)    indigenous
21801    0.2046570  0.0005737478
21809    0.2468368 -0.0214480873
40101    0.4328652 -0.2172053123
40102    0.2520883 -0.0880436406
40103    0.3374755 -0.1369644686
40104    0.1940756 -0.0193228193
40107    0.2229505 -0.0722629598
40109    0.2837605 -0.0206349563
...
```



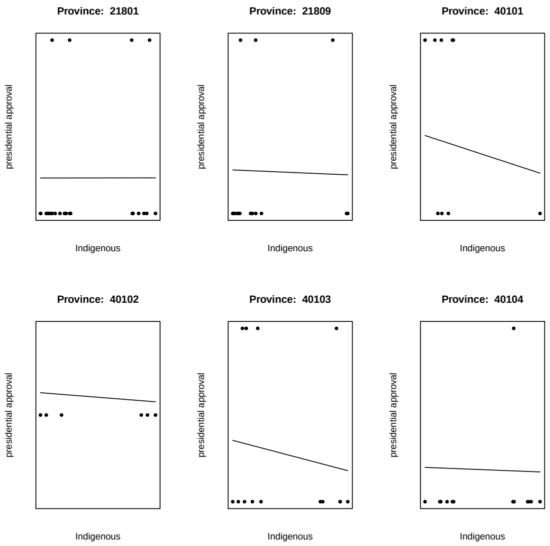
αj
βj
0.0
0.2
0.4
0.6
δ
−0.4
−0.2
0.0
0.2
0.4
λ
0
10
20
30
40
50
60
Province
Province

- As usual, we can also recover the "fixed effects", that give us the average presidential support and the average impact of *indigenous* on support, across all provinces

```
> display(M.4)
lmer(formula = president.approval ~ indigenous +
(1 + indigenous |     province))
             coef.est coef.se
(Intercept)   0.32     0.02
indigenous   -0.10     0.04

Error terms:
 Groups   Name        Std.Dev. Corr
 province (Intercept) 0.12
          indigenous  0.16     -0.64
 Residual             0.44
---
```

- We can also estimate the relationship $y_i = \alpha_j + \beta_j x_i$ for different js



Adding province-level predictors

- Finally, in M.5, we let the province-level predictor *illiteracy* affect α_j and β_j

$$y_i = \alpha_j + \beta_j x_i + \epsilon_i$$
$$\alpha_j = \delta + \gamma z_j + \nu_j$$
$$\beta_j = \lambda + \theta z_j + \eta_j$$

- Or:

$$y_i = \delta + \gamma z_j + \lambda x_i + \theta x_i z_j + \eta_j x_i + \epsilon_i + \nu_j$$

- Fixed effects: δ, γ, θ
- Level-2 random terms: $\eta_j(\times x_i)$, ν_j

- Estimating M.5 in R

```
M.5<-lmer(president.approval~indigenous+illiteracy+
illiteracy:indigenous+ (1+indigenous|province))
```

- Which leads to:

```
> fixef(M.5)
(Intercept) 0.6426787
indigenous -0.1947162
illiteracy  -4.1491954
indigenous:illiteracy 1.3739732
```

- And the estimated α_j and β_j:

```
coef(M.5)$province[,1:2]
        (Intercept)  indigenous
21801     0.5879079 -0.115671019
21809     0.5963156 -0.117241624
40101     0.6807045 -0.265152242
40102     0.6169996 -0.198411980
40103     0.6613938 -0.232940364
40104     0.5846869 -0.138587091
40107     0.6097006 -0.189449298
40109     0.6163529 -0.114249716
40110     0.6356596 -0.198797692
40112     0.5632993 -0.147592015
40122     0.6355085 -0.198642021
40129     0.6695299 -0.261848443
70101     0.5973700  0.006672967
70102     0.5787188 -0.107234567
...
```

Hierarchical generalized linear model

- So far, we have been working with a multilevel regression model
- However, *president.approval* is dichotomous: we need generalized linear models
- The main ideas seen so far apply to the hierarchical logit/probit models
- For instance, a hierarchical logit model using the specification in M.5 can be fit in R as:

```
M.6<-glmer(president.approval~indigenous+
illiteracy+
illiteracy:indigenous+
(1+indigenous|province), family=binomial(link="logit"))
```

```
>fixef(M.6)
      (Intercept)    0.7801372
 indigenous -0.4681024
illiteracy  -20.1687024
indigenous:illiteracy  -1.2554279

> coef(M.6)$province[,1:2]
       (Intercept)  indigenous
21801    0.6041166 -0.02203818
21809    0.6405997 -0.07144882
40101    0.8859700 -0.81265926
40102    0.7416090 -0.63746486
40103    0.8355677 -0.60797664
40104    0.6292574 -0.26798403
40107    0.7185312 -0.58788317
...
```

Only varying coefficient - not intercept

- So far, we have always assumed a random intercept: either alone, or together with a random coefficient
- However, in some circumstances, the researcher may believe that only the coefficient - not the intercept - is random
- This is not the convention, but nothing wrong with it
- This random coefficient model can be easily estimated in R: just exclude the constant from the term including the level-2 errors:

```
M.7<-glmer(president.approval~indigenous+
illiteracy+
illiteracy:indigenous+
(indigenous-1|province), family=binomial(link="logit"))
```

- Only β_j is random:

```
> coef(M.7)
$province
      (Intercept)  indigenous illiteracy indigenous:illiterac
21801   0.8081779 -0.37523944  -20.21813           -0.440046
21809   0.8081779 -0.38297492  -20.21813           -0.440046
40101   0.8081779 -0.72480237  -20.21813           -0.440046
40102   0.8081779 -0.83260434  -20.21813           -0.440046
40103   0.8081779 -0.60892002  -20.21813           -0.440046
40104   0.8081779 -0.61710467  -20.21813           -0.440046
40107   0.8081779 -0.82543782  -20.21813           -0.440046
40109   0.8081779 -0.15779440  -20.21813           -0.440046
40110   0.8081779 -0.65781732  -20.21813           -0.440046
40112   0.8081779 -0.94337438  -20.21813           -0.440046
40122   0.8081779 -0.65829325  -20.21813           -0.440046
40129   0.8081779 -0.80624976  -20.21813           -0.440046
70101   0.8081779  0.50709666  -20.21813           -0.440046
70102   0.8081779 -0.55495046  -20.21813           -0.440046
```

Non-nested hierarchies

- Multilevel models are also useful when we have non-nested structures; e.g., overlapping categories of attributes
- For example, suppose you are interested in analyzing earnings given occupation and region of residence
- That is, we might have survey data with individuals classified into
  - 40 job categories
  - 50 states
- Or, in the example of Peru, you could decide to treat ethnicity (white, indigenous, mestizo, other) as a "hierarchical" category alongside province
- How would you set up a multilevel model in this setting?

- Well, we could include random terms both for province and ethnicity.

- A simple random intercept model could look like this

$$
\begin{aligned}
y_i &= \delta_j + \gamma_k + \alpha + \beta x_i + \epsilon_i \\
\delta_j &= \eta_j \\
\gamma_k &= \nu_k
\end{aligned}
$$

or, equivalently (assuming normality of all the error terms)

$$
\begin{aligned}
y_i &\sim N(\delta_j + \gamma_k + \alpha + \beta x_i, \sigma_\epsilon^2) \\
\delta_j &\sim N(0, \sigma_\eta^2) \\
\gamma_k &\sim N(0, \sigma_\nu^2)
\end{aligned}
$$

where δ_j are province random effects, γ_k are ethnicity random effects, α and β are fixed effects, and x_i is a Level-1 (survey respondent) variable

How to fit this non-nested hierarchical model in R

```
M.8<-lmer(president.approval~age+
(1|province)+(1|ethnicity))
```

```
> ranef(M.8)

$province $ethnicity
21801  -0.0776007459 1 -0.048686942
21809  -0.0435204728 2 -0.038140526
40101   0.0882742533 3 -0.022006495
40102  -0.0695926788 4  0.022424173
40103   0.0002898258 5  0.037936870
40104  -0.0974322824 6  0.006730721
40107  -0.0990445242 7  0.019368851
40109  -0.0003647124 8  0.022373348
```

When R is not enough

- Commands like *lmer()* and *glmer()* in R are fine for the basic multi-level models we have seen today.
- However, multilevel models can become very complicated
  - many varying intercepts, slopes, non-nested components
  - several "hierarchies"
  - small samples that don't allow for an efficient estimation of σ_ν and σ_η
  - models that do not have "closed forms" (i.e., very complicated likelihoods)
- In these settings, R is not enough $\rightarrow$ move to WinBUGS
- WinBUGS estimates a wide array of models (multilevel or not) using Bayesian methods
- Check out http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

Multi-level models in WinBUGS

- Bayesian perspective on multi-level models: random effects are simply another parameter to be estimated
- So, we need a prior distribution for the random effect, and need to derive a posterior distribution
- Example: Vowles, Stevens and Katz (forthcoming)
- Analyze determinants of turnout using rolling cross sections
  - account for generational and election heterogeneity through random effects
  - variation of age-period-cohort models
- Basic model:

$$Pr(Y_{i,t,c} = 1) = X\beta + \eta_t + \nu_c \tag{3}$$

- Where:
  - β is a vector of fixed-effects
  - $\eta_t(0, \sigma_t^2)$ are election-specific random effects (prior)
  - $\nu_c(0, \sigma_c^2)$ are cohort-specific random effects (prior)

- What is the posterior distribution of η_t and ν_c?

$$Posterior \propto exp \frac{(Y^*_{i,t,c} - X\beta - \eta_t - \nu_c)^2}{} \prod_t ({\sigma^2}_t)^{-1/2} exp \frac{(\eta_t)^2}{\sigma_t^2} \prod_c ({\sigma^2}_c)^{-1/2} exp \frac{(\nu_c)^2}{\sigma_c^2} \tag{4}$$

$$\text{Posterior for } \eta_t \sim N\Big((N_{i,t} + \sigma_t^{-2})^{-1}(\sum_{i:i \in t} Y^*_{i,t,c} - X\beta - \nu_c); (N_{i,t} + \sigma_t^{-2})^{-1}\Big) \quad (5)$$

$$\text{Posterior for } \nu_c \sim N\Big((N_{i,c} + \sigma_c^{-2})^{-1}(\sum_{i:i \in t} Y^*_{i,t,c} - X\beta - \eta_t); (N_{i,c} + \sigma_c^{-2})^{-1}\Big) \quad (6)$$