STATISTICAL HORIZONS

Multilevel and Mixed Models Using R

Stephen Vaisey, Ph.D.

Upcoming Seminar:

July 30-31, 2019, Philadelphia, Pennsylvania

Multilevel and Mixed Models in Stata

Stephen Vaisey

1

About Me

- Professor of Sociology at Duke University
- I study statistical methods, culture, morality, politics, and social psychology

2

By tomorrow, you will…

- Get the "big picture" intuition of MLMs
- Know the difference between fixed and random effects
- Understand what random intercept models, random coefficient models, and crossed random effects models are and when to use each one
- Know how to estimate basic MLMs in Stata and interpret their results

3

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*

4

Why Use Multilevel Models?

- avoid problems with basic models (good)
- ask more interesting questions (better)

5

Sources for Example Datasets

http://www.stata-press.com/data/r14/me.html

net from http://www.stata-press.com/data/mlmus/
net get mlmus3_vol1
net get mlmus3_vol2

https://study.sagepub.com/robsonandpevalin

http://www.europeansocialsurvey.org/

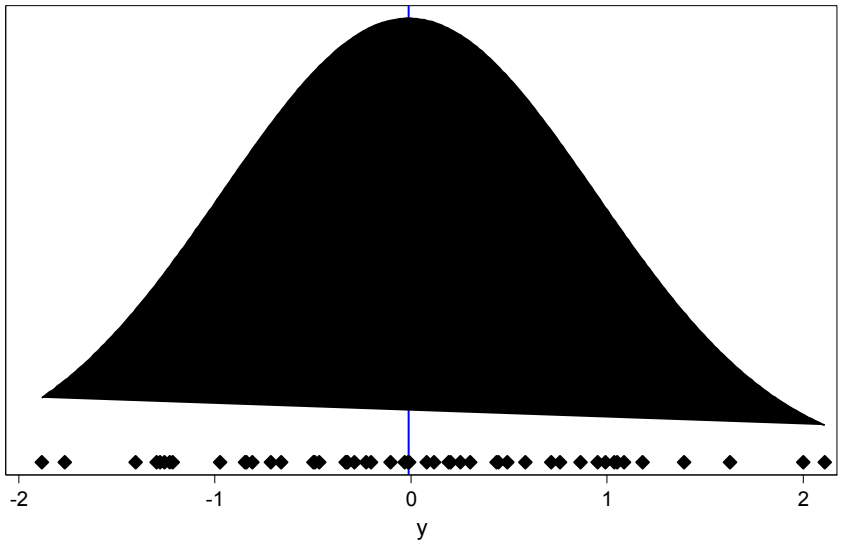
6

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

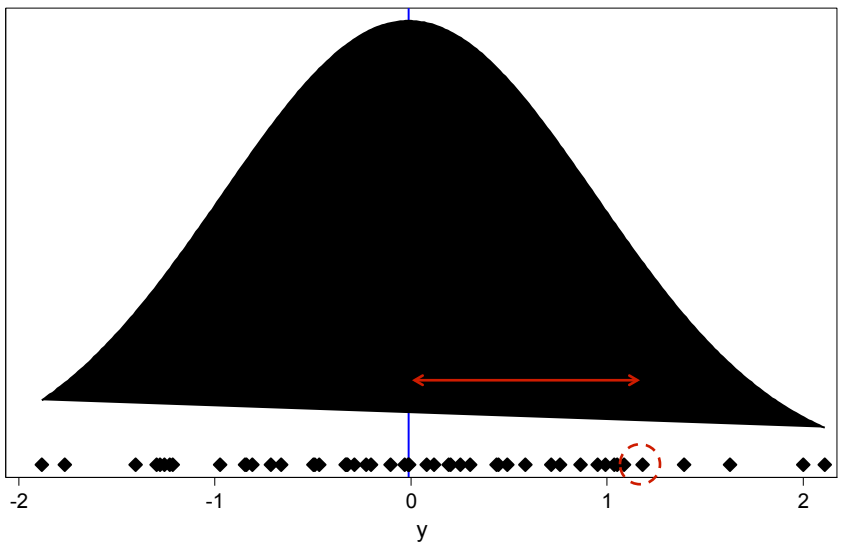
7

$$y_i = \mu + \varepsilon_i$$

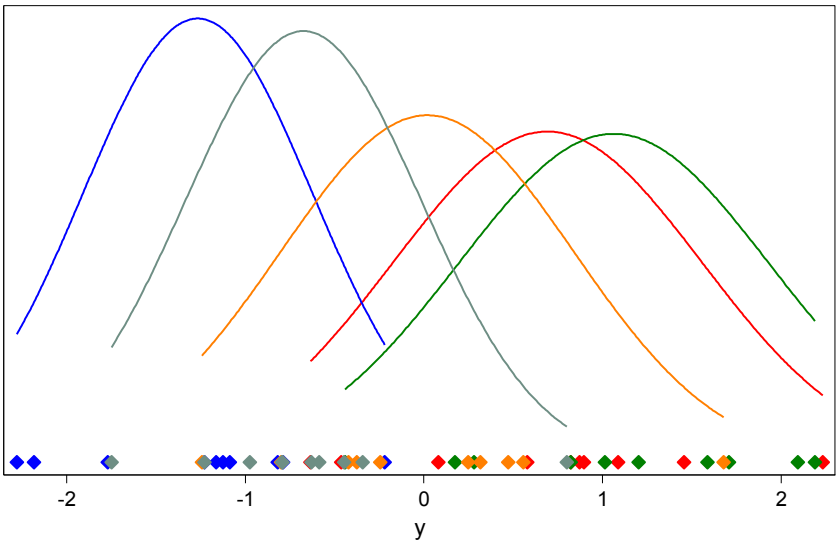
8



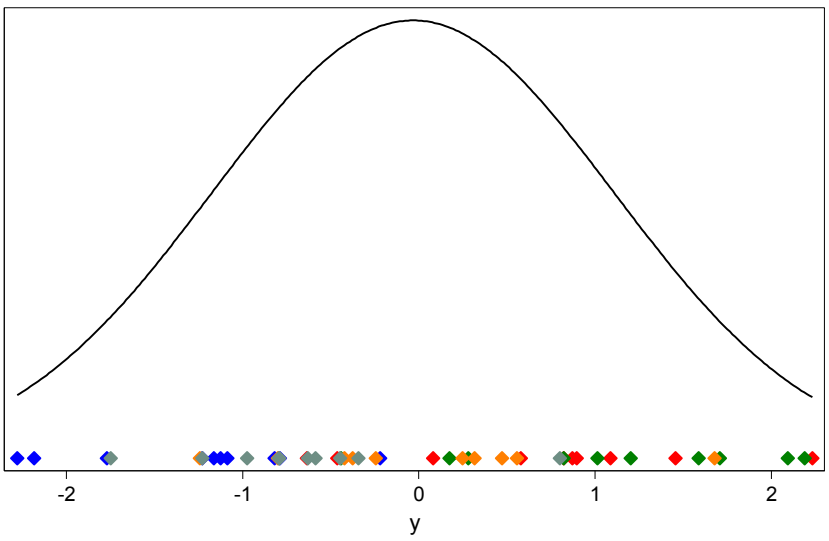
9



10



11

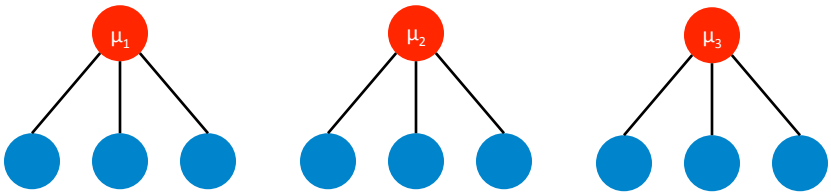


12

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

13

Hierarchical DGP



14

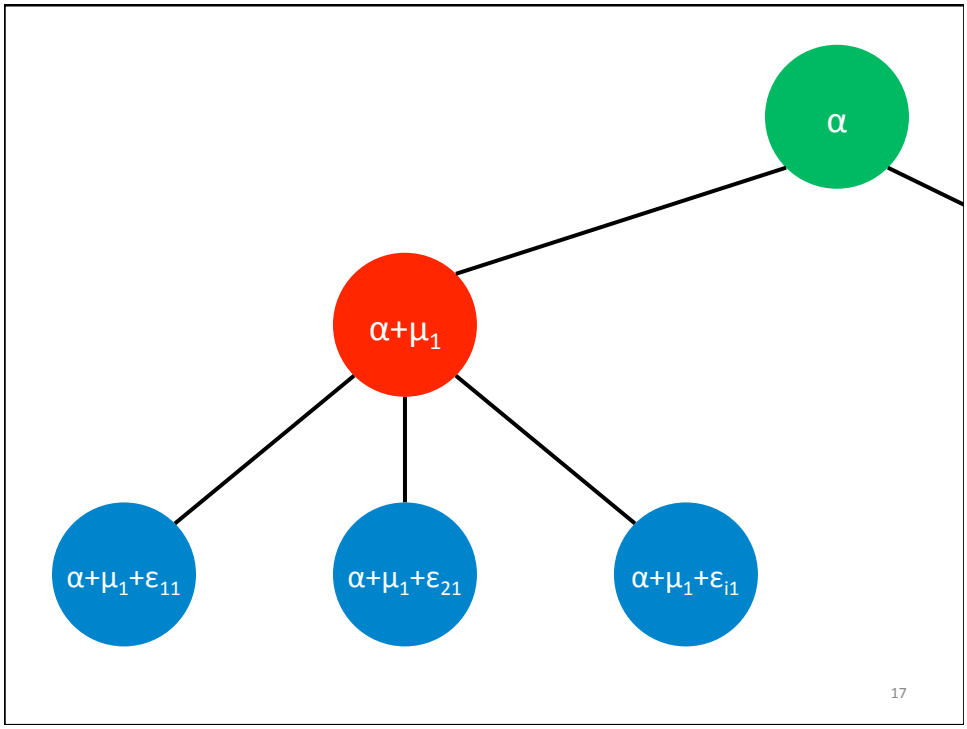
DGP Simulation Logic

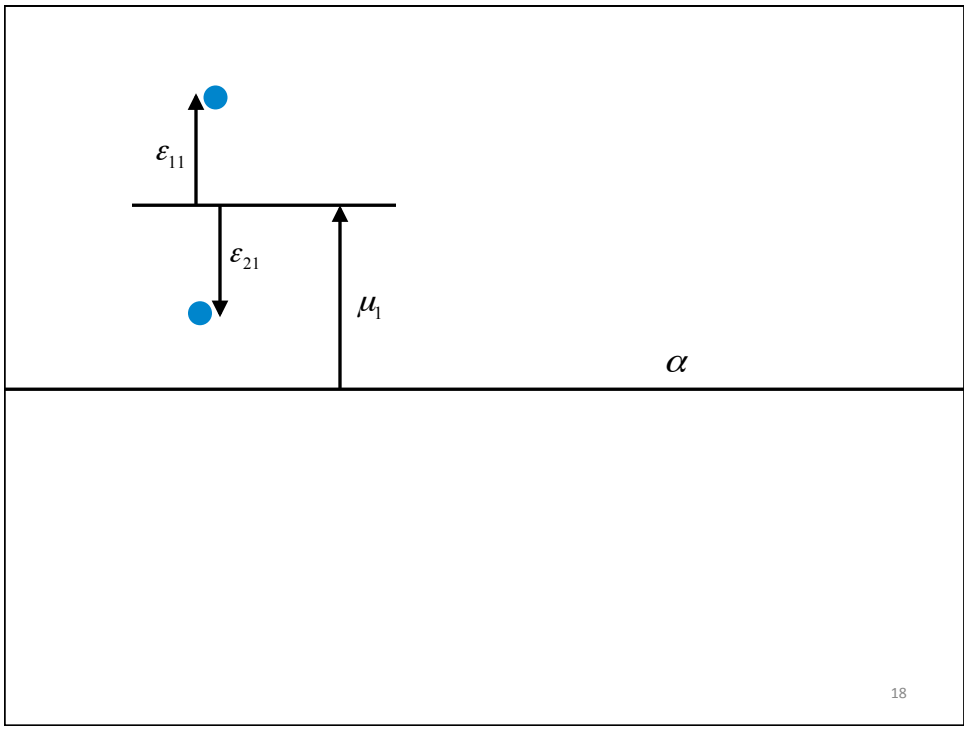
```
set obs 5                  // # of clusters
gen u = rnormal(0,1)       // the different μ values
expand 10                  // how many "people" per cluster
gen y = u + rnormal(0,1)   // observed variable
```

15

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

16


α
α+μ1
α+μ1+ε11
α+μ1+ε21
α+μ1+εi1
17


ε11
ε21
μ1
α
18

Different notations, same idea

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

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

$$y_{ij} = \beta + \zeta_j + \varepsilon_{ij}$$

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

19

Main Point

Because the observations are not independent, using only a single error term (ε_i) is not justified. The errors of the individual observations may be correlated, violating a core assumption of OLS regression.

However, if that's your *only* concern, that can be dealt with using clustered SEs.

20

Example:
Life Satisfaction in Europe

21

Data

200 person per country subsample of 2014 ESS (for faster estimation)

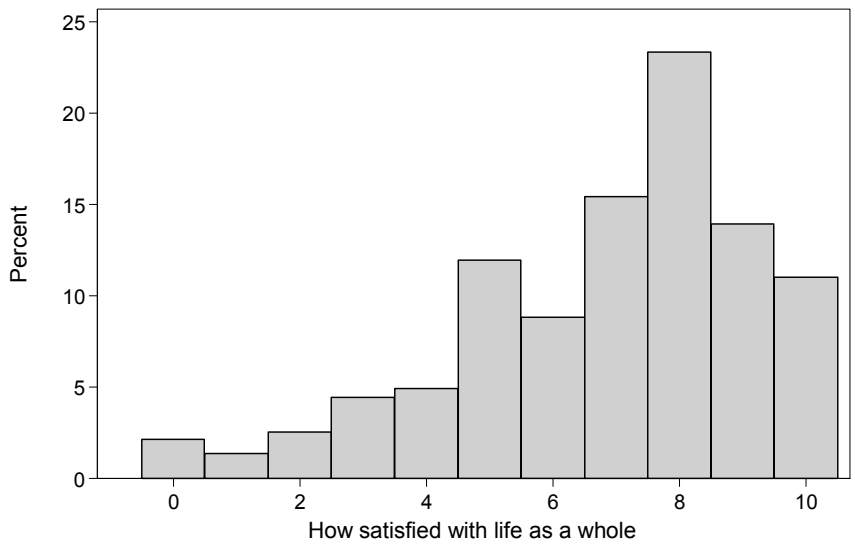
B20 **CARD 13** All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means extremely[18] dissatisfied and 10 means extremely satisfied.

Extremely dissatisfied										**Extremely satisfied**
00	01	02	03	04	05	06	07	08	09	10

22

How satisfied with life as a whole	Freq.	Percent	Cum.
Extremely dissatisfied	125	2.16	2.16
1	79	1.36	3.52
2	148	2.55	6.07
3	258	4.45	10.52
4	286	4.93	15.45
5	694	11.97	27.41
6	512	8.83	36.24
7	895	15.43	51.67
8	1,354	23.34	75.02
9	809	13.95	88.97
Extremely satisfied	640	11.03	100.00
Total	5,800	100.00	

23



24

```
. sum stflife, d
```

How satisfied with life as a whole

Percentiles		Smallest		
1%	0	0		
5%	2	0		
10%	3	0	Obs	5,800
25%	5	0	Sum of Wgt.	5,800
50%	7		Mean	6.783966
		Largest	Std. Dev.	2.414616
75%	8	10		
90%	10	10	Variance	5.830369
95%	10	10	Skewness	-.8581936
99%	10	10	Kurtosis	3.29755

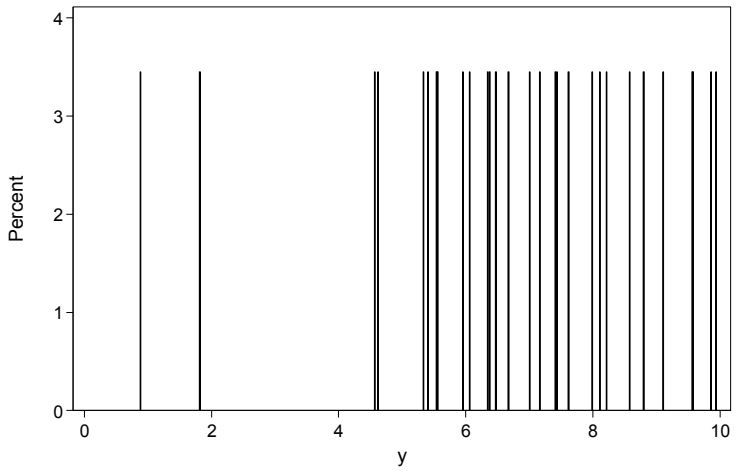
25

```
. tab country
```

Country	Freq.	Percent	Cum.
AL	200	3.45	3.45
BE	200	3.45	6.90
BG	200	3.45	10.34
CH	200	3.45	13.79
CY	200	3.45	17.24
CZ	200	3.45	20.69
DE	200	3.45	24.14
DK	200	3.45	27.59
EE	200	3.45	31.03
ES	200	3.45	34.48
FI	200	3.45	37.93
FR	200	3.45	41.38
GB	200	3.45	44.83
HU	200	3.45	48.28
IE	200	3.45	51.72
IL	200	3.45	55.17
IS	200	3.45	58.62
IT	200	3.45	62.07
LT	200	3.45	65.52
NL	200	3.45	68.97
NO	200	3.45	72.41
PL	200	3.45	75.86
PT	200	3.45	79.31
RU	200	3.45	82.76
SE	200	3.45	86.21
SI	200	3.45	89.66
SK	200	3.45	93.10
UA	200	3.45	96.55
XK	200	3.45	100.00
Total	5,800	100.00	

26

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



27

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

28