Longitudinal Data II

Roberta De Vito

BROWN
Public Health

Multi-level Model Q1

$$y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2)$$

$$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2)$$

Varying intercepts and slopes

$$y_i = N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2)$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right)$$

Varying intercepts and slopes: R output

```
M3 <- lmer (y ~ x + (1 + x | county))
display (M3)
```

which yields

```
lmer(formula = y ~ x + (1 + x | county))
            coef.est coef.se
(Intercept)  1.46     0.05
x           -0.68     0.09
Error terms:
 Groups   Name        Std.Dev. Corr
 county   (Intercept) 0.35
          x           0.34     -0.34

    Residual             0.75
   # of obs: 919, groups: county, 85
   deviance = 2161.1
```

Varying intercepts and slopes: R output

```
coef (M3)

$county
   (Intercept)     x
1         1.14 -0.54
2         0.93 -0.77
3         1.47 -0.67
. . .
85        1.38 -0.65
```

Varying intercepts and slopes: R output

```
fixef (M3)

(Intercept)         x
       1.46     -0.68
```

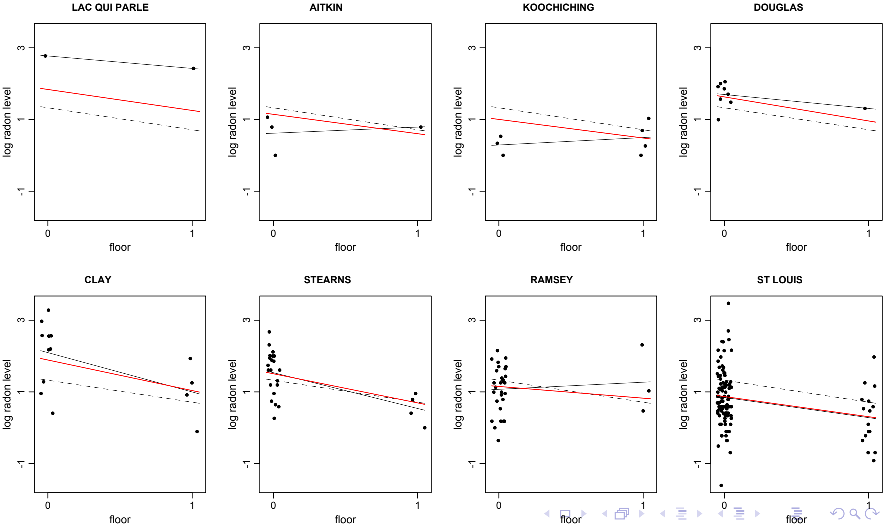
Varying intercepts and slopes: R output

```
ranef (M3)

(Intercept)        x
1           -0.32  0.14
2           -0.53 -0.09
3            0.01  0.01
. . .
85          -0.08  0.03
```

Multilevel regression lines $y = \alpha_j + \beta_j x$, displayed for eight counties j

Solid (no pool., $y = \alpha_j + \beta x$), Dashed (comp. pool., $y = \alpha + \beta x$)



Including group-level predictors: soil uranium

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2)$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_0^{\alpha} + \gamma_1^{\alpha} u_j \\ \gamma_0^{\beta} + \gamma_1^{\beta} u_j \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right)$$

Including group-level predictors: R output

Q2-Q3-Q4

```
lmer(formula = y ~ x + u.full + x:u.full + (1 + x | county))
            coef.est coef.se
(Intercept)  1.47     0.04
x           -0.67     0.08
u.full       0.81     0.09
x:u.full    -0.42     0.23
Error terms:
 Groups   Name        Std.Dev. Corr
 county   (Intercept) 0.12
          x           0.31     0.41
 Residual             0.75
# of obs: 919, groups: county, 85
deviance = 2114.3
```

Including group-level predictors: R output

The fixed effects

```
(Intercept)           x      u.full   x:u.full
       1.47       -0.67        0.81      -0.42
```

Including group-level predictors: R output

The group effects

```
   (Intercept)     x
1        -0.01  0.02
2         0.03 -0.21
. . .
85       -0.02 -0.03
```

Varying slopes as interactions

$$
\begin{aligned}
y_i &= \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i^y \\
\alpha_j &= \gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha \\
\beta_j &= \gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta
\end{aligned}
$$

Varying slopes as interactions

$$
\begin{aligned}
y_i &= \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i^y \\
\alpha_j &= \gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha \\
\beta_j &= \gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta
\end{aligned}
$$

$$\Downarrow$$

$$y_i = \left[\gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha\right] + \left[\gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta\right] x_i + \epsilon_i^y$$

Varying slopes as interactions

$$
\begin{aligned}
y_i &= \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i^y \\
\alpha_j &= \gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha \\
\beta_j &= \gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta
\end{aligned}
$$

$$\Downarrow$$

$$y_i = \left[\gamma_0^\alpha + \gamma_1^\alpha u_j + \epsilon_j^\alpha\right] + \left[\gamma_0^\beta + \gamma_1^\beta u_j + \epsilon_j^\beta\right] x_i + \epsilon_i^y$$

$$y_i = a + bv_i + c_{j[i]} + dx_i + ev_i x_i + f_{j[i]} x_i + \epsilon_i^y$$

Several ways

- A varying-intercept, varying-slope model with four individual-level predictors
- A regression model with $4 + 2J$ predictors
- A regression model with four predictors and three error terms
- A varying-intercept, varying-slope model with one group-level predictor

Constant-intercept, varying-slope model Q5

$$y_i \sim N(\alpha + \theta_{j[i]} T_i, \sigma_y^2)$$

$$\theta_j \sim N(\mu_\theta, \sigma_\theta^2)$$

Constant-intercept, varying-slope model: adding the x

$$y_i \sim N(\alpha + \beta x_i + \theta_{1,j[i]} T_i + \theta_{2,j[i]} x_i T_i, \sigma_y^2)$$

$$\begin{pmatrix} \theta_{1,j[i]} \\ \theta_{2,j[i]} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$$

Multilevel logistic regression

- Opinions at the levels of individual states as well as for the entire country
- Demographic factors such as sex, ethnicity, age, and education
- y_i: 1 Republican, 0 Democrat
- The state index is $j[i]$ with 51 states

Multilevel logistic regression

```
M1 <- lmer(y ~ black + female + (1|state), family=binomial(link="logit"))
display (M1)
```

Multilevel logistic regression

```
M1 <- lmer(y ~ black + female + (1|state), family=binomial(link="logit"))
display (M1)
```

```
              coef.est  coef.se
(Intercept)    0.4       0.1
black         -1.7       0.2
female        -0.1       0.1
Error terms:
 Groups         Name        Std.Dev.
 state          (Intercept) 0.4
 No residual sd
# of obs: 2015, groups: state, 49
deviance = 2658.7
  overdispersion parameter = 1.0
```

A fuller model: with 5 regions

$$Pr(y_i = 1) = logit^{-1}(\beta_0 + \beta^{female} female_i + \beta^{black} black_i + \beta^{female.black} female_i black_i + \alpha^{age}_{k[i]} + \alpha^{edu}_{l[i]} + \alpha^{age.edu}_{k[i]l[i]} + \alpha^{state}_{j[i]}$$

A fuller model: with 5 regions

$$
\begin{aligned}
Pr(y_i = 1) = & logit^{-1}(\beta_0 + \beta^{female} female_i + \beta^{black} black_i + \\
& \beta^{female.black} female_i black_i + \alpha^{age}_{k[i]} + \alpha^{edu}_{l[i]} + \alpha^{age.edu}_{k[i]l[i]} + \alpha^{state}_{j[i]}
\end{aligned}
$$

$$
\alpha^{state}_{j[i]} \sim N\left(\alpha^{region}_{m[j]} + \beta^{v.prev} v.prev_j, \sigma^2_{state}\right)
$$

A fuller model: with 5 regions

$$
\begin{aligned}
Pr(y_i = 1) = & logit^{-1}(\beta_0 + \beta^{female} female_i + \beta^{black} black_i + \\
& \beta^{female.black} female_i black_i + \alpha^{age}_{k[i]} + \alpha^{edu}_{l[i]} + \alpha^{age.edu}_{k[i]l[i]} + \alpha^{state}_{j[i]}
\end{aligned}
$$

$$
\alpha^{state}_{j[i]} \sim N\left(\alpha^{region}_{m[j]} + \beta^{v.prev} v.prev_j, \sigma^2_{state}\right)
$$

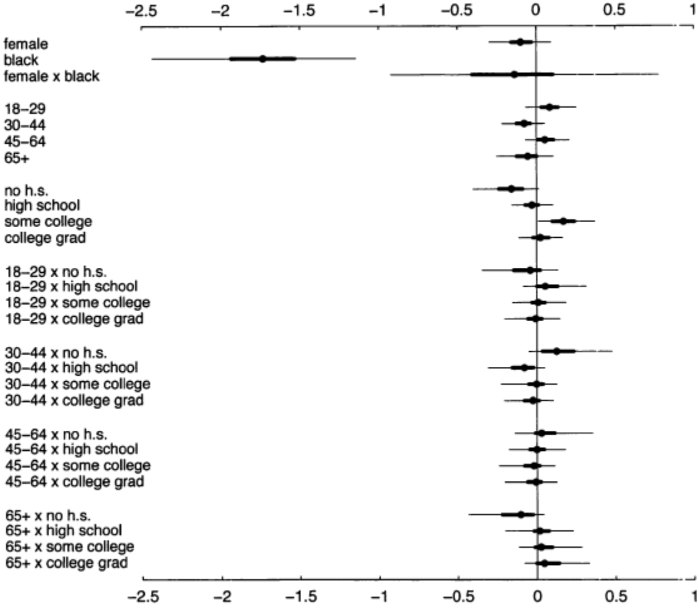
$$
\begin{aligned}
\alpha^{age}_{k} &\sim N(0, \sigma^2_{age}) \\
\alpha^{edu}_{l} &\sim N(0, \sigma^2_{edu}) \\
\alpha^{age.edu}_{k,l} &\sim N(0, \sigma^2_{age.edu}) \\
(\alpha^{region}_{m} &\sim N(0, \sigma^2_{region})
\end{aligned}
$$

A fuller model: output

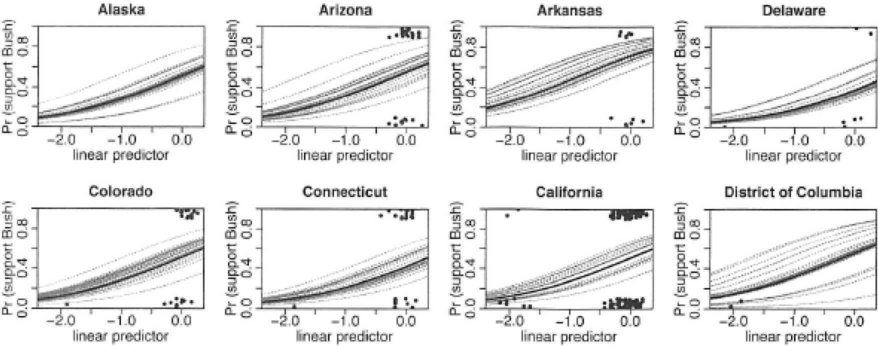
```
lmer(formula = y ~ black + female + black:female + v.prev.full +
    (1 | age) + (1 | edu) + (1 | age.edu) + (1 | state) +
    (1 | region.full), family = binomial(link = "logit"))
             coef.est coef.se
(Intercept)  -3.5      1.0
black        -1.6      0.3
female       -0.1      0.1
v.prev.full   7.0      1.7
black:female -0.2      0.4
Error terms:
 Groups        Name        Std.Dev.


 state         (Intercept) 0.2
 age.edu       (Intercept) 0.2
 region.full   (Intercept) 0.2
 edu           (Intercept) 0.1
 age           (Intercept) 0.0
 No residual sd
# of obs: 2015, groups: state,49; age.edu,16; region.full,5; edu,4; age,4
deviance = 2629.5
  overdispersion parameter = 1.0
```

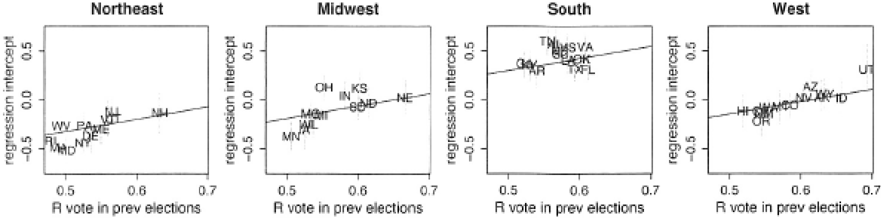
Graphing the estimated model



Graphing the estimated model

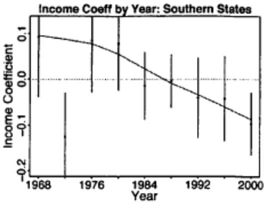


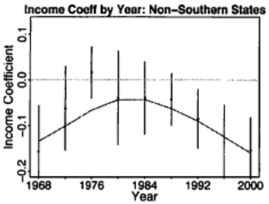
Graphing the estimated model



Over time

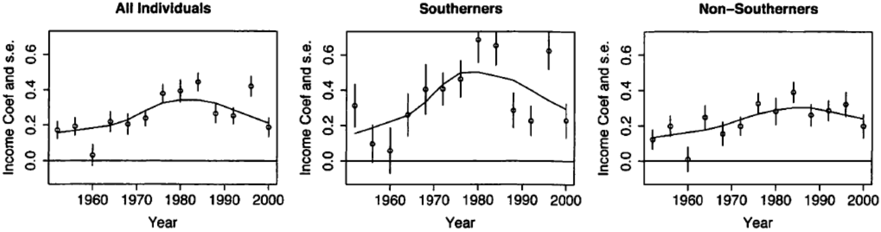
Richer States





Over time

National Election Study



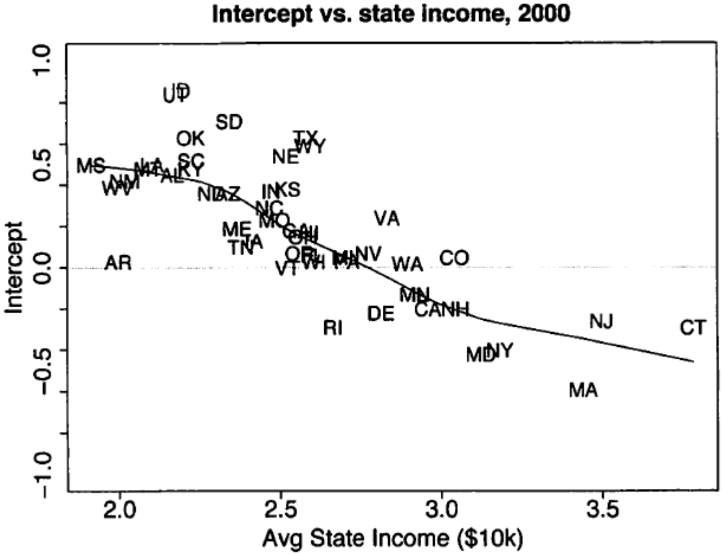
Varying-intercept model of income and vote preference within states

- 2000 presidential election using the National Annenberg Election Survey
- Fit a multilevel model

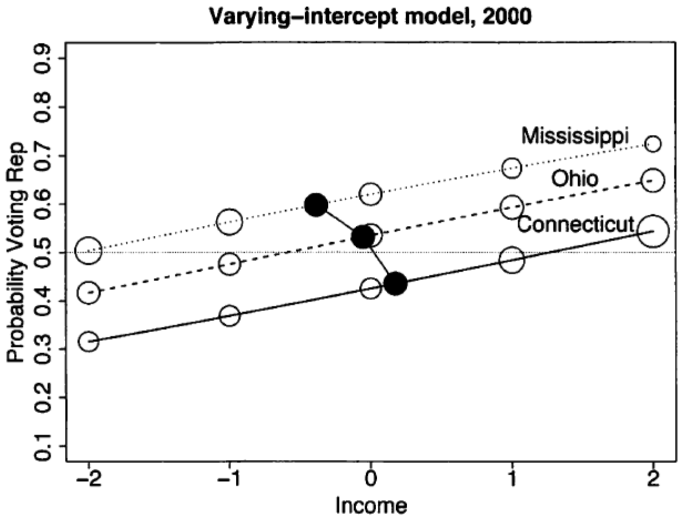
$$Pr(yi = 1) = logit^1(\alpha_{j[i]} + \beta x_i),$$

- $j[i]$: the state (from 1 to 50) corresponding to respondent i,
- x_i: persons household income (on the five-point scale),
- n: respondents in the poll
- $\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2)$
- u_j state average income level

Varying-intercept model of income and vote preference within states



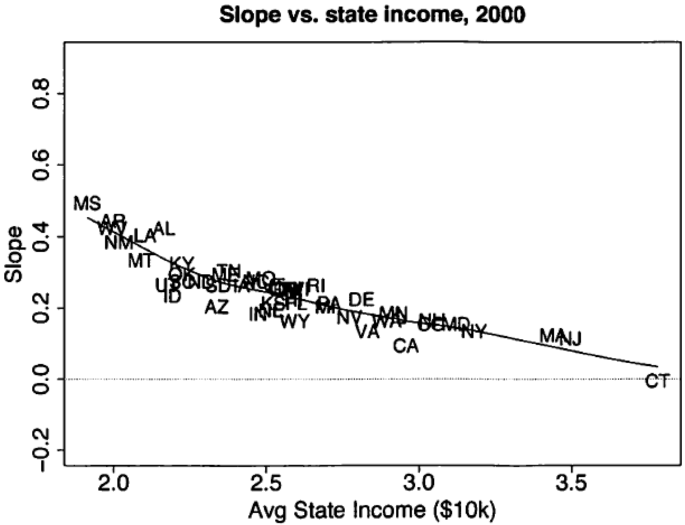
The paradox is not a paradox



Varying-intercept, varying-slope model

$$pr(y_i = 1) = logit^{-1}(\alpha_{j[i]} + \beta_{j[i]}x_i)$$
$$\alpha_j = \gamma_0^{\alpha} + \gamma_1^{\alpha} u_j + \epsilon_j^{\alpha}$$
$$\beta_j = \gamma_0^{\beta} + \gamma_1^{\beta} u_j + \epsilon_j^{\beta}$$

Varying-intercept, varying-slope model: the plot



Positive slopes within states and a negative slope between states