Motivating Multilevel Models

David Ryan Miller

February 13, 2020

Today's Goals

- Understand When and Why to Use Multilevel Models
- Know How to Fit and Interpret Multilevel Models

White House Visitor Logs

- Quarterly observations of interest groups (2010 Q1-2016 Q3)
- Sample data set structure
  - 11,155 observations
  - 500 unique interest groups
  - 27 quarters
- Key Variables
  - Number of visits (t)
  - Number of visits (t-1)
  - Lobbying expenditures ($100,000 units) (t-1)
  - Group CFscores

White House Visitor Logs: A Look

```r
head(final_df[which(final_df$IGName=="Amazon.com"),], n=10)
```

```
##     num_visits num_visits1l lobbyexp1l CFscore YearQuarter     IGName
##  1:          1            0    6.45001   0.069      2010Q1 Amazon.com
##  2:          0            1    6.15001   0.069      2010Q2 Amazon.com
##  3:          0            0    5.85001   0.069      2010Q3 Amazon.com
##  4:          1            0    5.55001   0.069      2010Q4 Amazon.com
##  5:          0            1    6.15001   0.069      2011Q1 Amazon.com
##  6:          3            0    7.35001   0.069      2011Q2 Amazon.com
##  7:          0            3    5.45001   0.069      2011Q3 Amazon.com
##  8:          2            0    5.45001   0.069      2011Q4 Amazon.com
##  9:          2            2    9.15001   0.069      2012Q1 Amazon.com
## 10:          3            2    8.65001   0.069      2012Q2 Amazon.com
```

Hypothesis

- More lobbying expenditures → More WH visits

```
ols_model <- lm(num_visits ~ num_visits1l + lobbyexp1l
                + as.factor(IGName)
                + as.factor(YearQuarter),
                data = final_df)
```

Visits ~ Lobbying Expenditures (OLS)

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	2.25	0.68	3.3	0
NumVisits1l	0.38	0.01	42.9	0
LobbyExp1l	0.07	0.01	5.9	0

Revisiting OLS Assumptions

- No perfect multicollinearity
- Variability in X
- Linearity
- Strict exogeneity ($\mathbf{E}[\epsilon_i|\mathbf{X}] = 0$)
- Constant variance of errors ($Var(\epsilon_i|\mathbf{X} = \sigma^2)$)
- Non-autocorrelation for errors ($Cov(\epsilon_i, \epsilon_j|\mathbf{X} = 0$ for $i \neq j$))

Revisiting OLS Assumptions

- No perfect multicollinearity
- Variability in X
- Linearity
- Strict exogeneity ($\mathbf{E}[\epsilon_i|\mathbf{X}] = 0$)
- Constant variance of errors ($Var(\epsilon_i|\mathbf{X} = \sigma^2)$)
- Non-autocorrelation for errors ($Cov(\epsilon_i, \epsilon_j|\mathbf{X} = 0$ for $i \neq j$))

Going Multilevel

- OLS Model
  - $visits_{i,j,p,t} = \zeta + \alpha_j + \gamma_p + \beta_1 visits_{i,t-1} + \beta_2 lobbyexp_{i,t-1}$ $+\epsilon_{i,j,p,t}$
- Multilevel linear model
  - $visits_{i,t} = \alpha_{j[i]} + \gamma_{p[t]} + \beta_1 visits_{i,t-1} + \beta_2 lobbyexp_{i,t-1} + \epsilon_{i,t}$
  - $\alpha_j \sim \mathsf{N}(\mu_\alpha, \sigma^2_\alpha)$
  - $\gamma_p \sim \mathsf{N}(\mu_\gamma, \sigma^2_\gamma)$

What Does Going Multilevel Buy Us?

- Accounts for heirarchical structure of data
  - Correlation among observations

Hypothesis Revisited

- More lobbying expenditures $\rightarrow$ More WH visits

```
lmer_model <- lmer(num_visits ~ num_visits1l + lobbyexp1l
                   + (1|IGName) + (1|YearQuarter),
                   data = final_df)
```

Visits $\sim$ Lobbying Expenditures (MLM)

	Estimate	Std. Error	t	Pr(>|t|)
(Intercept)	1.06	0.17	6.20	0.00
num_visits1l	0.46	0.01	54.63	0.00
lobbyexp1l	0.15	0.01	15.72	0.00
sd(IGName)	2.02			
sd(YearQuarter)	0.69			
sd(Residual)	3.42			

What Does Going Multilevel Buy Us?

- Accounts for heirarchical structure of data
  - Correlation among observations
- Modeling of group-level predictors
  - More efficient estimates of group-varying intercepts
  - Can include both group-varying intercepts and group-level predictors

Old and New Hypotheses

- More lobbying expenditures → More WH visits
- More liberal groups → More WH visits

```
lmer_model <- lmer(num_visits ~ num_visits1l + lobbyexp1l
                   + CFscore + (1|IGName)
                   + (1|YearQuarter), data = final_df)
```

Visits ~ Lobbying Expenditures + CFScores (MLM)

	Estimate	Std. Error	t	Pr(>|t|)
(Intercept)	1.20	0.18	6.67	0.00
num_visits1l	0.46	0.01	54.60	0.00
lobbyexp1l	0.15	0.01	15.76	0.00
CFscore	-0.62	0.25	-2.53	0.01
sd(IGName)	2.01			
sd(YearQuarter)	0.69			
sd(Residual)	3.42			

What Does Going Multilevel Buy Us?

- Accounts for heirarchical structure of data
  - Correlation among observations
- Modeling of group-level predictors
  - More efficient estimates of group-varying intercepts
  - Can include both group-varying intercepts and group-level predictors
- Modeling of group-varying slopes and variances
  - Account for heterogeneous effects
  - Account for heteroskedasticity

Additional Resources

- Applied Publications
  - Gelman and Hill (2007), *Data Analysis Using Regression and Multilevel/Heirarchical Models*
  - Shor, Bafumi, Keele, and Park (2007) "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data," *Political Analysis*
- `R` Packages
  - Maximum likelihood: `lme4`
  - Bayesian: `brms`, `rstan`, `rstanarm`