STA 610L: Module 4.1

Genealized linear mixed effects models (Part I)

Dr. Olanrewaju Michael Akande

STA 610L

Generalized linear mixed effects model (GLMM)

As we continue to generalize the concepts we have covered, let's think about the incorporation of random effects into the standard representation of generalized linear models.

The basic idea is that we assume there is natural heterogeneity across groups in a subset of the regression coefficients.

These coefficients are assumed to vary across groups according to some distribution.

Conditional on the random effects, we then assume the responses for a single subject are independent observations from a distribution in the exponential family.

GLMM

Note: when we look at longitudinal data, we will group by i, otherwise, we will group by j.

In the generalized linear mixed effects model (GLMM) for longitudinal data, we assume the conditional distribution of each Y_{ij}, conditional on $\mathbf{b}_i$, belongs to the exponential family with conditional mean

$$g(E[Y_{ij} \mid \mathbf{b}_i]) = \mathbf{X}_{ij}'\boldsymbol{\beta} + \mathbf{Z}_{ij}'\mathbf{b}_i,$$

where $g(\cdot)$ is a known link function.

Assume the $\mathbf{b}_i$ are independent across subjects with $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$.

We also assume that given $\mathbf{b}_i$, the responses $Y_{i1}, \ldots, Y_{in}$ are mutually independent.

Example: multilevel linear regression

$$Y_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij},$$

where

$$b_i \overset{iid}{\sim} N(0, \sigma_b^2) \perp \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$$

and

$$E(Y_{ij} \mid b_i) = \mathbf{X}'_{ij}\boldsymbol{\beta} + b_i$$

STA 610L

Example: multilevel logistic model with random intercepts

$$\text{logit}(E(Y_{ij} \mid b_i)) = \mathbf{X}_{ij}'\boldsymbol{\beta} + b_i,$$

where

$$b_i \sim N(0, \sigma^2)$$

Question: what happened to ε_{ij}?

Example: random coefficients Poisson regression

$$\log(E(Y_{ij} \mid \mathbf{b}_i)) = \mathbf{X}_{ij}'\boldsymbol{\beta} + \mathbf{Z}_{ij}'\mathbf{b}_i.$$

If we set

$$\mathbf{X}_{ij} = \mathbf{Z}_{ij} = [1, t_{ij}],$$

that is, we have random slopes and intercepts, then we can assume

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}).$$

Interpretation of GLMM estimates

In the model

$$\text{logit}(E[Y_{ij} \mid b_i]) = \mathbf{X}'_{ij}\boldsymbol{\beta} + b_i,$$

with $b_i \sim N(0, \sigma^2)$, each element of $\boldsymbol{\beta}$ measures the change in the log odds of a 'positive' response per unit change in the respective covariate, in a specific group that has an underlying propensity to respond positively given by b_i.

That is, we need to hold the group membership constant when interpreting β_k, just as we would hold the values of $\mathbf{x_{i,(-k)}}$ constant when interpreting β_k

STA 610L

CAUTION

Note also that with a non-linear link function, a non-linear contrast of the averages is not equal to the average of non-linear contrasts, so that the parameters do not in general have population-average interpretations (as they would in a linear mixed effects model, which has identity link).

So while in the lmm

$$g(E(Y_{ij} \mid \mathbf{X}_{ij}, \mathbf{b}_i)) = \mathbf{X}'_{ij}\boldsymbol{\beta} + \mathbf{Z}'_{ij}\mathbf{b}_i$$

so that $E(Y_{ij} \mid \mathbf{X}_{ij}) = \mathbf{X}'_{ij}\boldsymbol{\beta}$, when $g(\cdot)$ is non-linear (say the logit), then

$$g(E(Y_{ij} \mid \mathbf{X}_{ij})) \neq \mathbf{X}'_{ij}\boldsymbol{\beta}$$

for all $\boldsymbol{\beta}$ when averaged over the distribution of the random effects.

Intraclass correlation

Consider an unobserved continuous variable W_{ij}.

W_{ij} is related to Y_{ij} in the following manner: $Y_{ij} = 1$ if $W_{ij} < c$, and $Y_{ij} = 0$ otherwise.

The location of c and the distribution of W govern the probability that $Y = 1$.

Intraclass correlation

Useful way of thinking about model but not an essential assumption:

$$W_{ij} = \mathbf{X}'_{ij}\boldsymbol{\beta} + b_i + \varepsilon_{ij}$$

- $\varepsilon_{ij} \sim N(0,1)$: probit regression
- $\varepsilon_{ij} \sim$ standard logistic (mean 0, variance $\frac{\pi^2}{3}$): logistic regression

We can use this framework to calculate ICC's:

- $ICC = \frac{\sigma^2}{\sigma^2+1} \quad \text{for probit}$
- $ICC = \frac{\sigma^2}{\sigma^2+\frac{\pi^2}{3}} \quad \text{for logistic}$

Estimation using ML

The joint probability density function is given by

$$f(\mathbf{Y}_i \mid \mathbf{X}_i, \mathbf{b}_i) f(\mathbf{b}_i).$$

However, recall that the $\mathbf{b}_i$ are unobserved.

How then do we handle the $\mathbf{b}_i$ in the maximization?

Typically, we base frequentist inferences on the marginal (integrated) likelihood function, given by

$$\prod_{i=1}^{N} \int f(\mathbf{Y}_i \mid \mathbf{X}_i, \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i.$$

Estimation using maximum likelihood then involves a two-step procedure.

STA 610L

ML estimation steps

Step 1: Obtain ML estimates of $\boldsymbol{\beta}$ and $\mathbf{D}$ based on the marginal likelihood of the data.

While this may sound simple, numerical or Monte Carlo integration techniques are typically required, and the techniques used may have substantial impacts on resulting inferences.

Step 2: Given estimates of $\boldsymbol{\beta}$ and $\mathbf{D}$, predict the random effects as

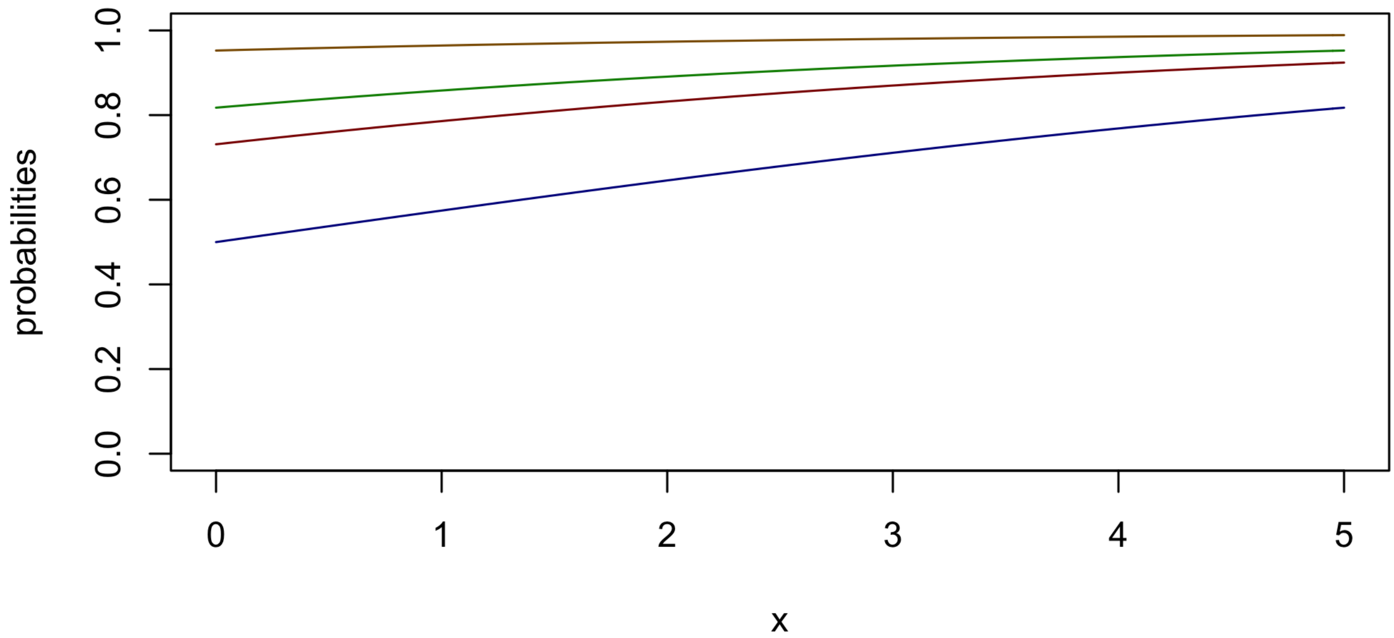
$$\widehat{\mathbf{b}}_i = E(\mathbf{b}_i \mid \mathbf{Y}_i, \widehat{\boldsymbol{\beta}}, \widehat{\mathbf{D}}).$$

Again, simple analytic solutions are rarely available.

Even when the burden of integration is not that great, the optimization problem may be difficult to solve.

Random effects logistic regression

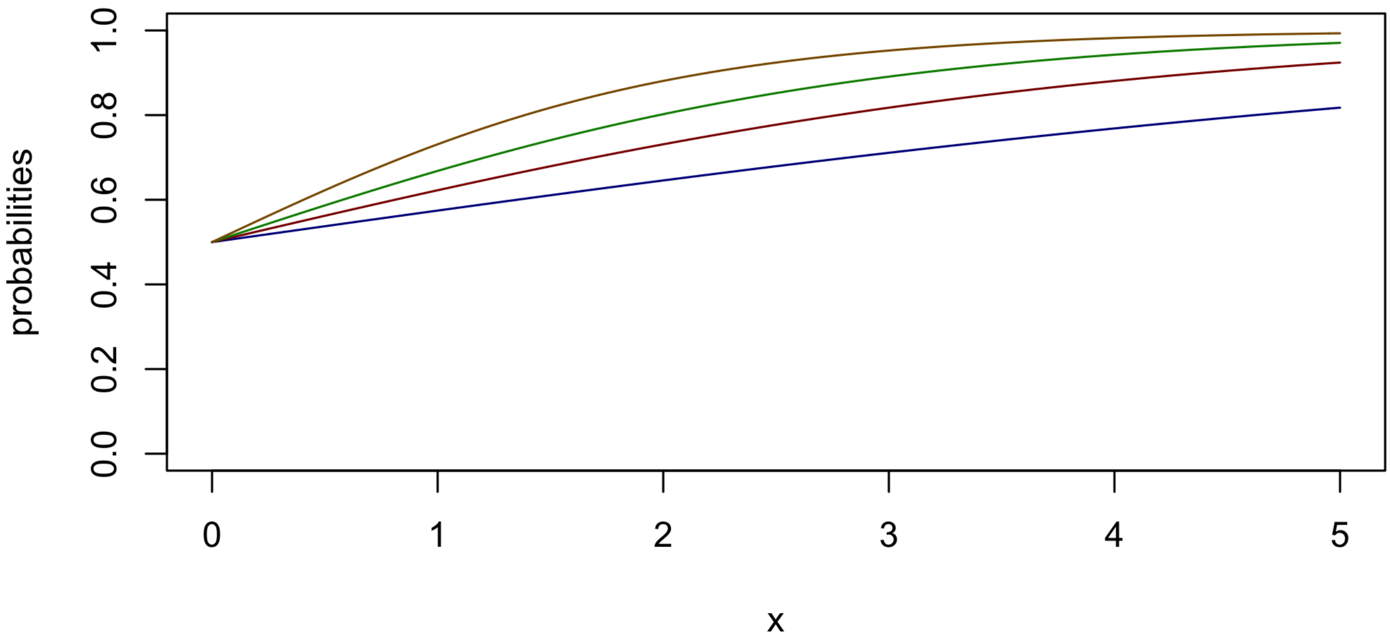
Inverse logit functions for random intercepts logistic model with a single predictor.



STA 610L

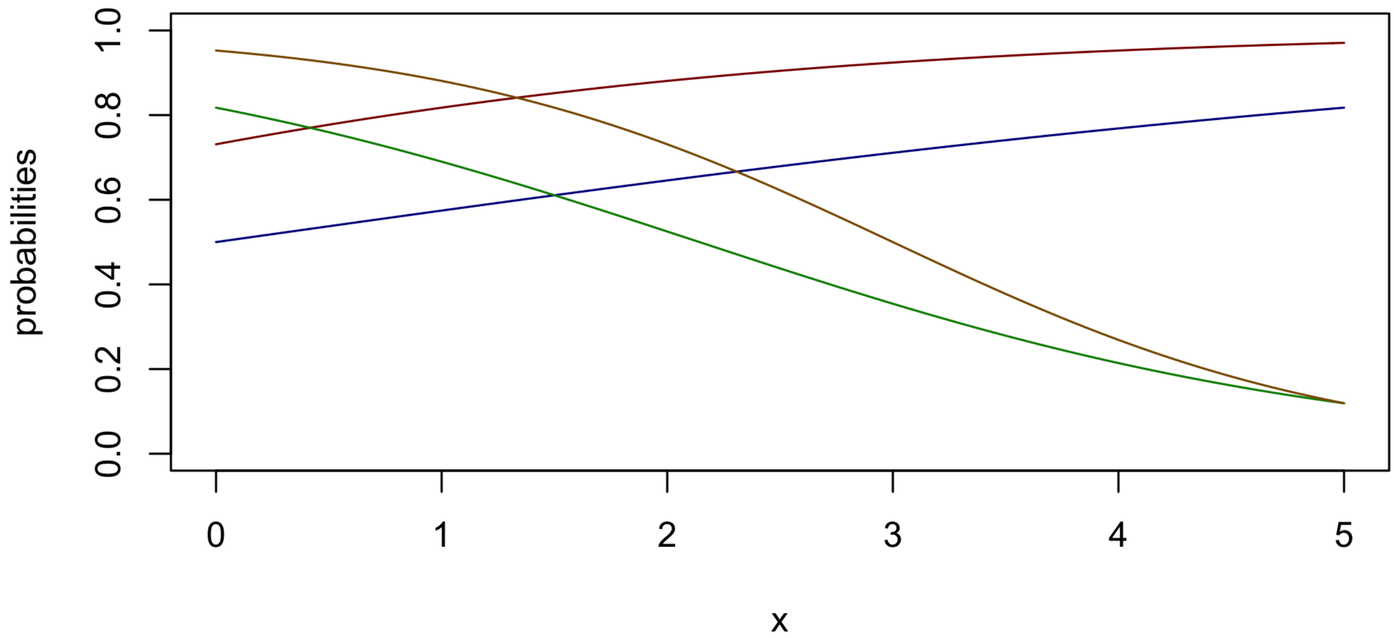
RANDOM EFFECTS LOGISTIC REGRESSION

Inverse logit functions for random slopes logistic model with a single predictor.



Random effects logistic regression

Inverse logit functions for random intercepts and random slopes logistic model with a single predictor.



1988 ELECTIONS ANALYSIS

To illustrate how to fit and interpret the results of random effect logistic models, we will use a sample data on election polls.

National opinion polls are conducted by a variety of organizations (e.g., media, polling organizations, campaigns) leading up to elections.

While many of the best opinion polls are conducted at a national level, it can also be often interesting to estimate voting opinions and preferences at the state or even local level.

Well-designed polls are generally based on national random samples with corrections for nonresponse based on a variety of demographic factors (e.g., sex, ethnicity, race, age, education).

The data is from CBS News surveys conducted during the week before the 1988 election.

Respondents were asked about their preferences for either the Republican candidate (Bush Sr.) or the Democratic candidate (Dukakis).

1988 ELECTIONS ANALYSIS

The dataset includes 2193 observations from one of eight surveys (the most recent CBS News survey right before the election) in the original full data.

Variable	Description
org	cbsnyt = CBS/NYT
bush	1 = preference for Bush Sr., 0 = otherwise
state	1-51: 50 states including DC (number 9)
edu	education: 1=No HS, 2=HS, 3=Some College, 4=College Grad
age	1=18-29, 2=30-44, 3=45-64, 4=65+
female	1=female, 0=male
black	1=black, 0=otherwise
region	1=NE, 2=S, 3=N, 4=W, 5=DC
v_prev	average Republican vote share in the three previous elections (adjusted for home-state and home-region effects in the previous elections)

Given that the data has a natural multilevel structure (through `state` and `region`), it makes sense to explore hierarchical models for this data.

STA 610L

1988 ELECTIONS ANALYSIS

Both voting turnout and preferences often depend on a complex combination of demographic factors.

In our example dataset, we have demographic factors such as biological sex, race, age, education, which we may all want to look at by state, resulting in $2 \times 2 \times 4 \times 4 \times 51 = 3264$ potential categories of respondents.

We may even want to control for `region`, adding to the number of categories.

Clearly, without a very large survey (most political survey poll around 1000 people), we will need to make assumptions in order to even obtain estimates in each category.

We usually cannot include all interactions; we should therefore select those to explore (through EDA and background knowledge).

The data is in the file `polls_subset.txt` on Sakai.

1988 ELECTIONS ANALYSIS

```r
###### Load the data
polls_subset <- read.table("data/polls_subset.txt",header=TRUE)
str(polls_subset)
```

```
## 'data.frame':    2193 obs. of  10 variables:
##  $ org   : chr  "cbsnyt" "cbsnyt" "cbsnyt" "cbsnyt" ...
##  $ survey: int  9158 9158 9158 9158 9158 9158 9158 9158 9158 9158 ...
##  $ bush  : int  NA 1 0 0 1 1 1 1 0 0 ...
##  $ state : int  7 39 31 7 33 33 39 20 33 40 ...
##  $ edu   : int  3 4 2 3 2 4 2 2 4 1 ...
##  $ age   : int  1 2 4 1 2 4 2 4 3 3 ...
##  $ female: int  1 1 1 1 1 1 0 1 0 0 ...
##  $ black : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ region: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ v_prev: num  0.567 0.527 0.564 0.567 0.524 ...
```

```r
head(polls_subset)
```

```
##      org survey bush state edu age female black region    v_prev
## 1 cbsnyt   9158   NA     7   3   1      1     0      1 0.5666333
## 2 cbsnyt   9158    1    39   4   2      1     0      1 0.5265667
## 3 cbsnyt   9158    0    31   2   4      1     0      1 0.5641667
## 4 cbsnyt   9158    0     7   3   1      1     0      1 0.5666333
## 5 cbsnyt   9158    1    33   2   2      1     0      1 0.5243666
## 6 cbsnyt   9158    1    33   4   4      1     0      1 0.5243666
```

STA 610L

1988 ELECTIONS ANALYSIS

```
summary(polls_subset)
```

```
##      org                survey          bush            state      
##  Length:2193        Min.   :9158   Min.   :0.0000   Min.   : 1.00  
##  Class :character   1st Qu.:9158   1st Qu.:0.0000   1st Qu.:14.00  
##  Mode  :character   Median :9158   Median :1.0000   Median :26.00  
##                     Mean   :9158   Mean   :0.5578   Mean   :26.11  
##                     3rd Qu.:9158   3rd Qu.:1.0000   3rd Qu.:39.00  
##                     Max.   :9158   Max.   :1.0000   Max.   :51.00  
##                                    NA's   :178                     
##       edu             age            female           black        
##  Min.   :1.000   Min.   :1.000   Min.   :0.0000   Min.   :0.00000  
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:0.0000   1st Qu.:0.00000  
##  Median :2.000   Median :2.000   Median :1.0000   Median :0.00000  
##  Mean   :2.653   Mean   :2.289   Mean   :0.5887   Mean   :0.07615  
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:0.00000  
##  Max.   :4.000   Max.   :4.000   Max.   :1.0000   Max.   :1.00000  
##                                                                    
##      region          v_prev      
##  Min.   :1.000   Min.   :0.1530  
##  1st Qu.:2.000   1st Qu.:0.5278  
##  Median :2.000   Median :0.5481  
##  Mean   :2.431   Mean   :0.5550  
##  3rd Qu.:3.000   3rd Qu.:0.5830  
##  Max.   :5.000   Max.   :0.6927  
## 
```

1988 ELECTIONS ANALYSIS

```
polls_subset$v_prev <- polls_subset$v_prev*100 #rescale
polls_subset$region_label <- factor(polls_subset$region,levels=1:5,
                                    labels=c("NE","S","N","W","DC"))
#we consider DC as a separate region due to its distinctive voting patterns
polls_subset$edu_label <- factor(polls_subset$edu,levels=1:4,
                                 labels=c("No HS","HS","Some College","College Grad"))
polls_subset$age_label <- factor(polls_subset$age,levels=1:4,
                                 labels=c("18-29","30-44","45-64","65+"))
#the data includes states but without the names, which we will need,
#so let's grab that from R datasets
data(state)
#"state" is an R data file (type ?state from the R command window for info)
state.abb #does not include DC, so we will create ours
```

```
##  [1] "AL" "AK" "AZ" "AR" "CA" "CO" "CT" "DE" "FL" "GA" "HI" "ID" "IL" "IN" "IA"
## [16] "KS" "KY" "LA" "ME" "MD" "MA" "MI" "MN" "MS" "MO" "MT" "NE" "NV" "NH" "NJ"
## [31] "NM" "NY" "NC" "ND" "OH" "OK" "OR" "PA" "RI" "SC" "SD" "TN" "TX" "UT" "VT"
## [46] "VA" "WA" "WV" "WI" "WY"
```

```
#In the polls data, DC is the 9th "state" in alphabetical order
state_abbr <- c (state.abb[1:8], "DC", state.abb[9:50])
polls_subset$state_label <- factor(polls_subset$state,levels=1:51,labels=state_abbr)
rm(list = ls(pattern = "state")) #remove unnecessary values in the environment
```

1988 ELECTIONS ANALYSIS

```
###### View properties of the data
head(polls_subset)
```

```
##      org survey bush state edu age female black region   v_prev region_label
## 1 cbsnyt   9158   NA     7   3   1      1     0      1 56.66333           NE
## 2 cbsnyt   9158    1    39   4   2      1     0      1 52.65667           NE
## 3 cbsnyt   9158    0    31   2   4      1     0      1 56.41667           NE
## 4 cbsnyt   9158    0     7   3   1      1     0      1 56.66333           NE
## 5 cbsnyt   9158    1    33   2   2      1     0      1 52.43666           NE
## 6 cbsnyt   9158    1    33   4   4      1     0      1 52.43666           NE
##      edu_label age_label state_label
## 1 Some College     18-29          CT
## 2 College Grad     30-44          PA
## 3           HS       65+          NJ
## 4 Some College     18-29          CT
## 5           HS     30-44          NY
## 6 College Grad       65+          NY
```

```
dim(polls_subset)
```

```
## [1] 2193   14
```

STA 610L

1988 ELECTIONS ANALYSIS

```r
###### View properties of the data
str(polls_subset)
```

```
## 'data.frame':    2193 obs. of  14 variables:
##  $ org         : chr  "cbsnyt" "cbsnyt" "cbsnyt" "cbsnyt" ...
##  $ survey      : int  9158 9158 9158 9158 9158 9158 9158 9158 9158 9158 ...
##  $ bush        : int  NA 1 0 0 1 1 1 1 0 0 ...
##  $ state       : int  7 39 31 7 33 33 39 20 33 40 ...
##  $ edu         : int  3 4 2 3 2 4 2 2 4 1 ...
##  $ age         : int  1 2 4 1 2 4 2 4 3 3 ...
##  $ female      : int  1 1 1 1 1 1 0 1 0 0 ...
##  $ black       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ region      : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ v_prev      : num  56.7 52.7 56.4 56.7 52.4 ...
##  $ region_label: Factor w/ 5 levels "NE","S","N","W",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ edu_label   : Factor w/ 4 levels "No HS","HS","Some College",..: 3 4 2 3 2 4 2 2 4 1 ...
##  $ age_label   : Factor w/ 4 levels "18-29","30-44",..: 1 2 4 1 2 4 2 4 3 3 ...
##  $ state_label : Factor w/ 51 levels "AL","AK","AZ",..: 7 39 31 7 33 33 39 20 33 40 ...
```

STA 610L

1988 ELECTIONS ANALYSIS

I will not do any meaningful EDA here.

I expect you to be able to do this yourself.

Let's just take a look at the amount of data we have for "bush" and the age:edu interaction.

```
###### Exploratory data analysis
table(polls_subset$bush) #well split by the two values
```

```
##
##    0    1
##  891 1124
```

```
table(polls_subset$edu,polls_subset$age)
```

```
##
##        1   2   3   4
##   1  44  42  67  96
##   2 232 283 223 116
##   3 141 205  99  54
##   4 119 285 125  62
```

STA 610L

1988 ELECTIONS ANALYSIS

As a start, we will consider a simple model with fixed effects of race and sex and a random effect for state (50 states + the District of Columbia).

$$
\begin{aligned}
\text{bush}_{ij} | \boldsymbol{x}_{ij} &\sim \text{Bernoulli}(\pi_{ij}); \quad i = 1, \ldots, n; \quad j = 1, \ldots, J = 51; \\
\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \beta_0 + b_{0j} + \beta_1 \text{female}_{ij} + \beta_2 \text{black}_{ij}; \\
b_{0j} &\sim N(0, \sigma^2).
\end{aligned}
$$

In `R`, we have

```
#library(lme4)
model1 <- glmer(bush ~ black+female+(1|state_label),
                family=binomial(link="logit"),
                data=polls_subset)
summary(model1)
```

1988 ELECTIONS ANALYSIS

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: bush ~ black + female + (1 | state_label)
##    Data: polls_subset
## 
##      AIC      BIC   logLik deviance df.resid 
##   2666.7   2689.1  -1329.3   2658.7     2011 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.7276 -1.0871  0.6673  0.8422  2.5271 
## 
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  state_label (Intercept) 0.1692   0.4113  
## Number of obs: 2015, groups:  state_label, 49
## 
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)  0.44523    0.10139   4.391 1.13e-05
## black       -1.74161    0.20954  -8.312  < 2e-16
## female      -0.09705    0.09511  -1.020    0.308
## 
## Correlation of Fixed Effects:
##        (Intr) black 
## black  -0.119       
## female -0.551 -0.005
```

STA 610L

1988 ELECTIONS ANALYSIS

Looks like we dropped some NAs.

```
c(sum(complete.cases(polls_subset)),sum(!complete.cases(polls_subset)))
```

```
## [1] 2015  178
```

Not ideal; we'll learn about methods for dealing with missing data soon.

Interpretation of results:

- For a fixed state (or across all states), a non-black male respondent has odds of $e^{0.45} = 1.57$ of supporting Bush.

- For a fixed state and sex, a black respondent as $e^{-1.74} = 0.18$ times (an 82% decrease) the odds of supporting Bush as a non-black respondent; you are much less likely to support Bush if your race is black compared to being non-black.

- For a given state and race, a female respondent has $e^{-0.10} = 0.91$ (a 9% decrease) times the odds of supporting Bush as a male respondent. However, this effect is not actually statistically significant!

1988 ELECTIONS ANALYSIS

The state-level standard deviation is estimated at 0.41, so that the states do vary some, but not so much.

I expect that you will be able to interpret the corresponding confidence intervals.

```
## Computing profile confidence intervals ...

##                  2.5 %      97.5 %
## .sig01       0.2608567  0.60403428
## (Intercept)  0.2452467  0.64871247
## black       -2.1666001 -1.34322366
## female      -0.2837100  0.08919986
```

1988 ELECTIONS ANALYSIS

We can definitely fit a more sophisticated model that includes other relevant survey factors, such as

- region
- prior vote history (note that this is a state-level predictor),
- age, education, and the interaction between them.

Given the structure of the data, it makes sense to include region as a second (nested) grouping variable.

We are yet to discuss that, so I will return to this later.

STA 610L

1988 ELECTIONS ANALYSIS

For now, let's just fit two models, one with the main effects for age and education, and the second with the interaction between them.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: bush ~ black + female + edu_label + age_label + (1 | state_label)
##    Data: polls_subset
## 
##      AIC      BIC   logLik deviance df.resid 
##   2662.2   2718.3  -1321.1   2642.2     2005 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.8921 -1.0606  0.6420  0.8368  2.7906 
## 
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  state_label (Intercept) 0.1738   0.4168  
## Number of obs: 2015, groups:  state_label, 49
## 
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.31206    0.19438   1.605  0.10841
## black                -1.74378    0.21124  -8.255  < 2e-16
## female               -0.09681    0.09593  -1.009  0.31289
## edu_labelHS           0.23282    0.16569   1.405  0.15998
## edu_labelSome College  0.51598    0.17921   2.879  0.00399
## edu_labelCollege Grad  0.31585    0.17454   1.810  0.07036
## age_label30-44       -0.29222    0.12352  -2.366  0.01800
## age_label45-64       -0.06744    0.13738  -0.491  0.62352
## age_label65+         -0.22509    0.16142  -1.394  0.16318
```

Can you interpret the results?

STA 610L

1988 ELECTIONS ANALYSIS

```
model3 <- glmer(bush ~ black + female + edu_label*age_label + (1|state_label),
                family=binomial(link="logit"),data=polls_subset)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00802313 (tol = 0.002, component 1)
```

Why do we have a rank deficient model? Also, it looks like we have a convergence issue.

These issues can happen. We have so many parameters to estimate from the interaction terms `edu_label*age_label` (16 actually), and it looks like that's causing a problem.

STA 610L

Note on estimation

ML estimation is carried out typically using adaptive Gaussian quadrature.

To improve accuracy over many package defaults (Laplace approximation), increase the number of quadrature points to be greater than one.

Note that some software packages require Laplace approximation with Gaussian quadrature if the number of random effects is more than 1 for the sake of computational efficiency.

What's next?

Move on to the readings for the next module!