Multilevel Models II

David M. Rocke

June 8, 2017

Radon Data Set

Individual home radon levels in the US, Minnesota subset of 919 households in 85 counties. All variables below are of length 919.

Variable	Definition
radon	Radon level in individual home
log.radon	Log-radon or 0.1 if radon=0
floor	$0 =$ basement, $1 =$ first floor
county.name	Name of each of county
county	County number, 1–85
ur	Uranium level in the county

County Level Variables

- We can add county-level variables, in this case a uranium level for each county.
- This is repeated for every household in the county.
- The county intercepts are then partly determined by the uranium content of the earth in that county.
- The remainder of the variability of the county levels is in the random intercept.

```
> summary(radon.lmer3)
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ floor + ur + (1 | county)

REML criterion at convergence: 2134.2

Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.02446  0.1564
 Residual             0.57523  0.7584
Number of obs: 919, groups:  county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.46576    0.03794   38.63
floor       -0.66824    0.06880   -9.71
ur           0.72027    0.09176    7.85

Correlation of Fixed Effects:
      (Intr) floor
floor -0.357
ur     0.145 -0.009
```

```
> summary(lmer(log.radon~floor+(1|county)))
Linear mixed model fit by REML ['lmerMod']
Formula: log.radon ~ floor + (1 | county)

REML criterion at convergence: 2171.3

Random effects:
 Groups    Name        Variance Std.Dev.
 county   (Intercept) 0.1077   0.3282
 Residual              0.5709   0.7556
Number of obs: 919, groups:  county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.46160    0.05158  28.339
floor       -0.69299    0.07043  -9.839

Correlation of Fixed Effects:
      (Intr)
floor -0.288
```

```
> summary(radon.lmer3)
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ floor + ur + (1 | county)

Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.02446  0.1564    #Note much smaller intercept variance
 Residual             0.57523  0.7584    #Due to uranium variable
Number of obs: 919, groups: county, 85


> summary(lmer(log.radon~floor+(1|county)))
Linear mixed model fit by REML ['lmerMod']
Formula: log.radon ~ floor + (1 | county)

Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.1077   0.3282
 Residual             0.5709   0.7556
Number of obs: 919, groups: county, 85
```

Group Level Variables

- Counties vary in their average radon level.
- Some of that variation may be due to identifiable, measurable variables such as uranium content in the earth.
- The remainder can be considered random.
- In this case, as generally, random means unexplained by other known factors or variables.

Multilevel Logistic Regression

- This is similar to multilevel regression except that the outcome is 0/1.
- We use `glmer` with `family=binomial`.
- Fixed effects coefficients are log odds ratios as in ordinary logistic regression.
- Random effects standard deviations are also on the log odds ratio scale, remembering that the mean value is 0.
- If the standard deviation is 0.4, then the random coefficient could vary from -0.8 to 0.8 on the log odds ratio scale, which is 0.45 to 2.23 on the odds ratio scale.

Gelman and Hill 1988 Election Polling Data

This is a national poll of 2193 individuals in the 48 contiguous states and DC taken before the 1988 election between George H.W. Bush (R) and Michael Dukakis (D). The 178 NAs are not counted in the logistic regression. Individual variables considered included race (1 = black, 0 = other), sex (1 = female, 0 = male), age (1–4, by category), and education (1–4, by category). We consider state as the primary grouping variable, with region as a group variable (1–5 = NE, S, N-Central, West, DC)

Simple Analysis

```
> summary(polls.glmer1)
Generalized linear mixed model fit by maximum likelihood
    (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: bush ~ black + female + (1 | state)
   Data: polls.subset

     AIC       BIC    logLik  deviance  df.resid
  2666.7    2689.1   -1329.3    2658.7      2011

Random effects:
 Groups Name        Variance Std.Dev.
 state  (Intercept) 0.1692   0.4113
Number of obs: 2015, groups:  state, 49
```

Odds ratio of voting for Bush compared to an average state varies from
about 0.44 to about 2.28 (using exp of 2 times coefficient).

We can extract the estimated values with ranef(polls.glmer1)$state
They vary from -0.557 to 0.648, or 0.57 to 1.91 on the odd ratio scale.
This is a smaller range because these estimates are shrunk to 0.

Simple Analysis

```
> summary(polls.glmer1)
Formula: bush ~ black + female + (1 | state)
   Data: polls.subset

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


These estimates have the usual logistic regression interpretation.
Large negative effect on preferring Bush from blacks, vs. non-blacks.
No significant effects of sex.
```

More Complex Analysis

```
> summary(polls.glmer2)
Generalized linear mixed model fit by maximum likelihood
    (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: bush ~ black * female + (1 | age) + (1 | edu) + (1 | age:edu) +
     (1 | state) + (1 | region.full)
   Data: polls.subset

     AIC      BIC   logLik deviance df.resid
  2662.4   2712.9  -1322.2   2644.4     2006

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8173 -1.0529  0.6450  0.8535  2.8997
```

More Complex Analysis

```
> summary(polls.glmer2)
Formula: bush ~ black * female + (1 | age) + (1 | edu) + (1 | age:edu) +
     (1 | state) + (1 | region.full)

Random effects:
 Groups       Name        Variance Std.Dev.
 state        (Intercept) 0.061386 0.24776
 age:edu      (Intercept) 0.007743 0.08799
 region.full (Intercept) 0.089055 0.29842
 edu          (Intercept) 0.020599 0.14352
 age          (Intercept) 0.007249 0.08514
Number of obs: 2015, groups:  state, 49; age:edu, 16; region.full, 5;
   edu, 4; age, 4
```

```
Individual-level variables and one state-level variable.
Largest effects are region and state, and then age and education
```

More Complex Analysis

```
> summary(polls.glmer2)
Formula: bush ~ black * female + (1 | age) + (1 | edu) + (1 | age:edu) +
    (1 | state) + (1 | region.full)

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.39374    0.19658   2.003   0.0452 *
black       -1.68200    0.32730  -5.139 2.76e-07 ***
female      -0.08946    0.09816  -0.911   0.3621
black:female -0.18200   0.42067  -0.433   0.6653
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation of Fixed Effects:
            (Intr) black  female
black       -0.086
female      -0.297  0.180
black:femal  0.059 -0.764 -0.232
```