A Multilevel Analysis of Radon Exposure in Minnesota

Valentina Litang

I built multilevel models to identify which counties should be prioritized for inspection. First, I created the following models and compared them to determine which model has the best fit.

- Model 1 - a basic model with log.u and floor as the fixed effects and county as the random effect
- Model 2 - same as model 1 but I added an interaction between log.u and floor
- Model 3 - same as model 1 but I added a random slope for log.u within county
- Model 4 - same as model 1 but I added a random slope for floor within county

AIC prefers model 4, and BIC prefers model 1. A likelihood ratio test between the two shows that model 4 does not significantly improve the model fit over model 1. Therefore, I will be using model 1.

```
radon <- read_csv("radon.csv")
```

```
## Rows: 919 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (6): log.u, uranium, radon, log.radon, floor, county
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
model1 <- lmer(log.radon ~ log.u + floor + (1|county), data = radon)
model2 <- lmer(log.radon ~ log.u * floor + (1|county), data = radon)
model3 <- lmer(log.radon ~ log.u + floor + (log.u | county), data = radon)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
model4 <- lmer(log.radon ~ log.u + floor + (floor | county), data = radon)

BIC(model1, model2, model3, model4)
```

```
##        df      BIC
## model1  5 2168.303
## model2  6 2171.845
## model3  7 2183.754
## model4  7 2176.403
```

```
AIC(model1, model2, model3, model4)
```

```
##        df      AIC
## model1  5 2144.186
## model2  6 2142.906
## model3  7 2149.991
## model4  7 2142.640
```

```
anova(model1, model4)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: radon
## Models:
## model1: log.radon ~ log.u + floor + (1 | county)
## model4: log.radon ~ log.u + floor + (floor | county)
##        npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
## model1    5 2132.8 2156.9 -1061.4    2122.8
## model4    7 2131.6 2165.4 -1058.8    2117.6 5.2139  2    0.07376 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next, I used my model to identify which counties should be prioritized for inspection.

```
random_effects <- ranef(model1)$county

random_effects <- data.frame(county = rownames(random_effects), effect = random_effects[, 1])

counties <- random_effects %>%
  arrange(desc(effect))

head(counties, 5)
```

```
##   county    effect
## 1      7 0.1281094
## 2     24 0.1280848
## 3     38 0.1260113
## 4     25 0.1253422
## 5     81 0.1218448
```

The five counties with the highest random effect should be prioritized. I found that counties 7, 24, 38, 25, and 81 should be prioritized.