

Multilevel Analysis 1(level1:individual)

CHEN MAOSEN

2025-12-19

Description of dataset

log.radon	basement	uranium	county	county.name
0.7885	1	-0.6890	1	AITKIN
0.7885	0	-0.6890	1	AITKIN
1.0647	0	-0.6890	1	AITKIN
0.0000	0	-0.6890	1	AITKIN
1.1314	0	-0.8473	2	ANOKA

Dataset: [Radon data](#)

- ▶ House variables:
 - ▶ **log.radon**: Numeric. Logarithm of indoor radon measurement values[min:-2.303/max:3.875/mean:1.225]
 - ▶ **basement**: Binary. Measurement position indication (0=basement/1=1st floor)
- ▶ County variables:
 - ▶ **county**: Factor. ID of each county with 85 levels
 - ▶ **county.name**: Factor. Name of each county
 - ▶ **uranium**: Numeric. Average uranium in the soil of each county[min:-0.882/max:0.528/mean:-0.132]

Analysis method and model specification

Random intercept model:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}\text{Basement}_{ij} + R_{ij} \quad R_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$\text{Level 2: } \beta_{0j} = \gamma_1 + U_j \quad U_j \sim \mathcal{N}(0, \tau_0^2)$$

$$\text{Combined: } Y_{ij} = \gamma_1 + \beta_{1j}\text{Basement}_{ij} + U_j + R_{ij}$$

$$Y_{ij} = \log.\text{radon}_{ij}$$

Parameters:

- ▶ γ_1 : Fixed effect. Mean intercept across all counties, controlling for Basement.
- ▶ β_{1j} : Fixed effect. Mean Basement slope across all counties.
- ▶ σ^2 : Random effect. Variance of county intercepts around γ_1 , controlling for Basement.
- ▶ τ_0^2 : Random effect. Variance of the houses around their county mean, controlling for Basement.

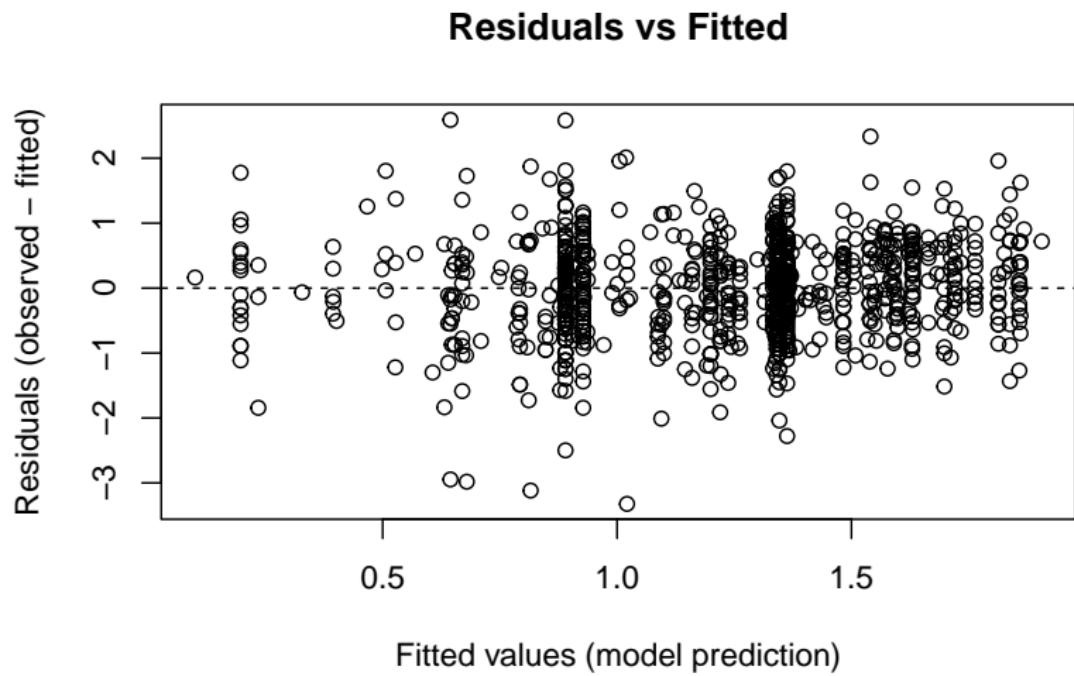
Analysis method and model specification

```
library(lme4)
Basement.fixed <- lmer(
  log.radon ~ 1 + basement + (1 | county),
  data = radon)

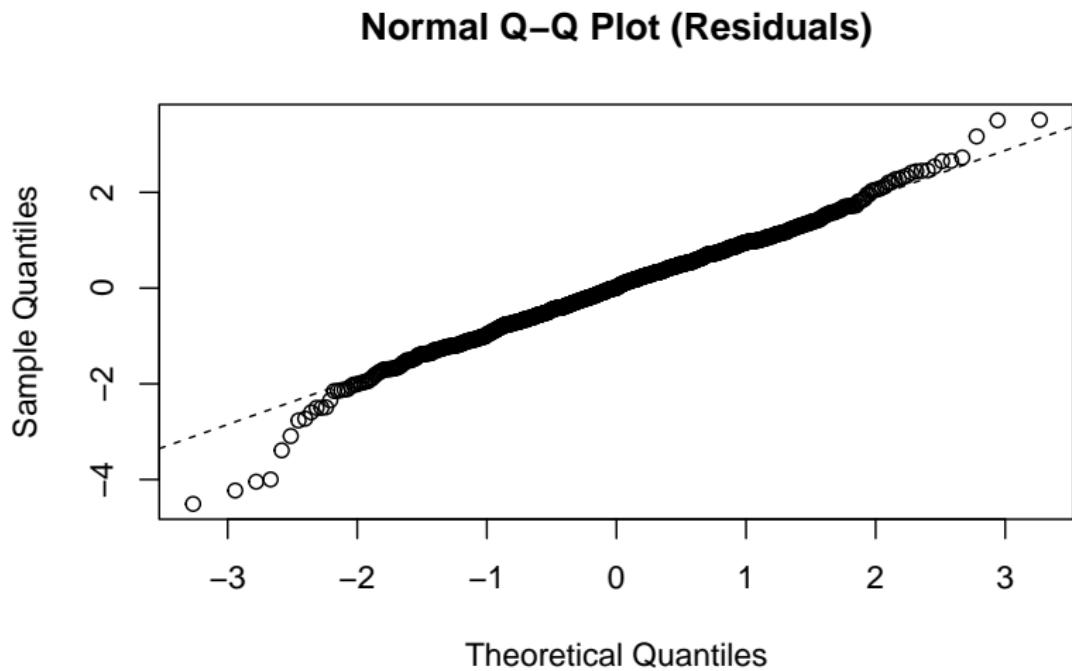
## Random effects:
##   Groups    Name        Variance Std.Dev.
##   county    (Intercept) 0.1077   0.3282
##   Residual             0.5709   0.7556
## Number of obs: 919, groups: county, 85
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)  1.46160   0.05158 28.339
## basement    -0.69299   0.07043 -9.839

## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.159 [0.085, 0.240]
##     Unadjusted ICC: 0.145 [0.076, 0.218]
```

Results

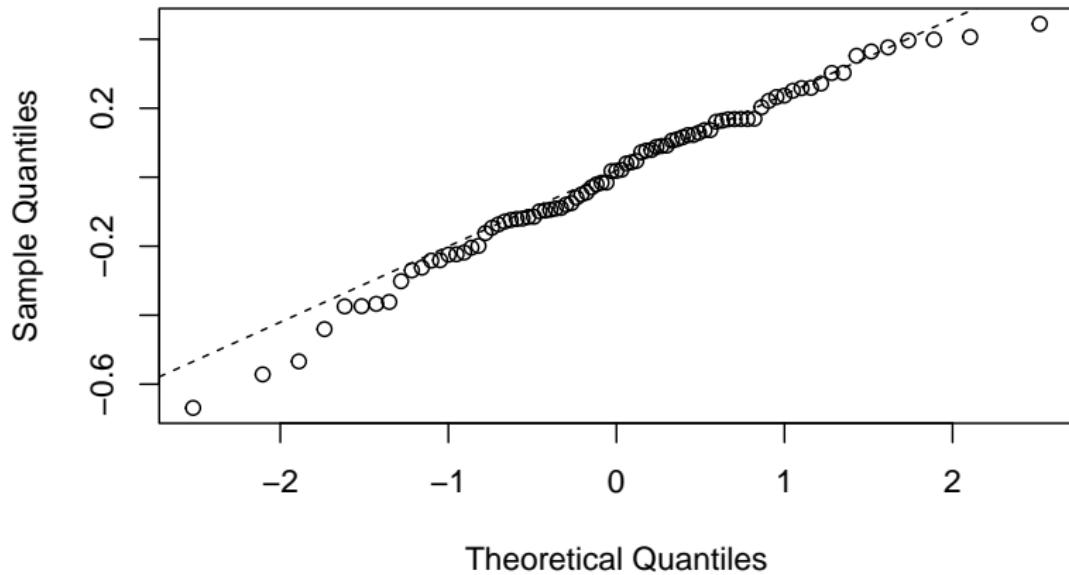


Results



Results

Normal Q-Q Plot (Random intercepts by county)



Results

Random intercept model:

By default, we use REML to estimate parameters:

$$\log.\text{radon}_{ij} = 1.462 - 0.693\text{Basement}_{ij} + U_j + R_{ij}$$

$$U_j \sim \mathcal{N}(0, 0.108)$$

$$R_{ij} \sim \mathcal{N}(0, 0.571)$$

ICC(95% CI)

Adjusted ICC: 0.159 [0.085, 0.240]

Unadjusted ICC: 0.145 [0.076, 0.218]

Interpretation:

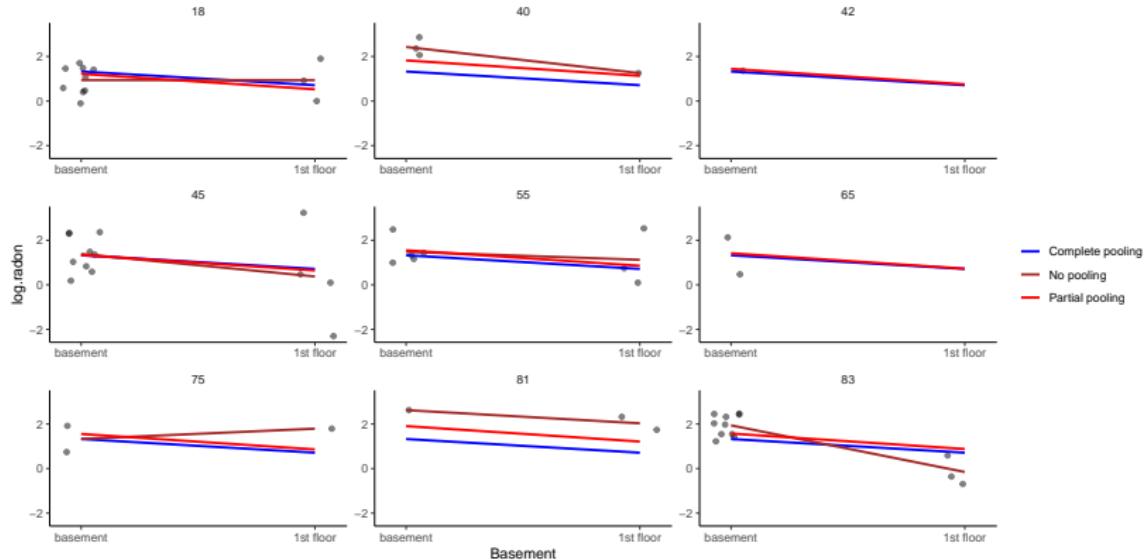
- ▶ Adjusted:

County grouping accounts for 15.9% of the total variance of the log.radon under the control of Basement.

- ▶ Unadjusted:

County grouping accounts for 14.5% of the total variance of the log.radon

Pooling



- ▶ Complete pooling assumes no variance between counties, but it's often larger.
- ▶ No pooling assumes larger variance between counties(independence), but it's often smaller.
- ▶ Partial pooling is in between.

Model comparison

Different fixed part (with level 2 predictor)

Model	Equation		Deviance	df
\mathcal{M}_0	$Y_{ij} = \gamma_1 + \beta_{1j}\text{Basement}_{ij}$	+	$U_j + R_{ij}$	$D_0 = 2163.7$
\mathcal{M}_1	$Y_{ij} = \gamma_1 + \beta_{1j}\text{Basement}_{ij} + \gamma_2\text{Uranium}_j$	+	$U_j + R_{ij}$	$D_1 = 2122.8$

Since models only have difference of fixed part, thus we **must** use ML to estimate parameters:

```
Basement.fixed      <- lmer(log.radon ~ 1 + basement +          (1 | county), data = radon)
Uranium.Basement.fixed <- lmer(log.radon ~ 1 + basement + uranium + (1 | county), data = radon)
anova(Basement.fixed,Uranium.Basement.fixed,
      refit=TRUE)
```

```
## Data: radon
## Models:
## Basement.fixed: log.radon ~ 1 + basement + (1 | county)
## Uranium.Basement.fixed: log.radon ~ 1 + basement + uranium + (1 | county)
##                      npar    AIC    BIC  logLik -2*log(L) Chisq Df
## Basement.fixed        4 2171.7 2190.9 -1081.8    2163.7
## Uranium.Basement.fixed 5 2132.8 2156.9 -1061.4    2122.8 40.834  1
##                      Pr(>Chisq)
## Basement.fixed
## Uranium.Basement.fixed 1.658e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model comparison

Different fixed part (with level 2 predictor)

Model	Equation	Deviance	df
\mathcal{M}_0	$Y_{ij} = \gamma_1 + \beta_{1j}\text{Basement}_{ij}$	$D_0 = 2163.7$	$v_0 = 4$
\mathcal{M}_1	$Y_{ij} = \gamma_1 + \beta_{1j}\text{Basement}_{ij} + \gamma_2\text{Uranium}_j$	$D_1 = 2122.8$	$v_1 = 5$

- ▶ Test statistic: $D_0 - D_1 \sim \chi^2_1$
- ▶ $D_0 - D_1 = 2163.7 - 2122.8 = 40.834$
- ▶ $p = P[D_0 - D_1 > 40.834] < .001$

Conclusion

At $\alpha = 5\%$, we reject \mathcal{M}_0 .

Adding Uranium significantly improves the model fit.

Model comparison

Different random part (with random slope)

Model	Equation	Deviance	df
M_0	$Y_{ij} = \gamma_1 + \gamma_2 \text{Uranium}_j + \beta_{1j} \text{Basement}_{ij} + U_{0j} + R_{ij}$	$D_0 = 2134.2$	$v_0 = 5$
M_1	$Y_{ij} = \gamma_1 + \gamma_2 \text{Uranium}_j + \beta_{1j} \text{Basement}_{ij} + U_{0j} + U_{1j} \text{Basement}_{ij} + R_{ij}$	$D_1 = 2128.6$	$v_1 = 7$

Since models only have difference of random part, thus we **should** use REML to estimate parameters:

```
Uranium.Basement.fixed <- lmer(log.radon ~ 1 + basement + uranium + (1 | county), data = radon)
Uranium.Basement.random <- lmer(log.radon ~ 1 + basement + uranium + (1 + basement|county), data = radon)
anova(Uranium.Basement.fixed, Uranium.Basement.random,
      refit=FALSE)
```

```
## Data: radon
## Models:
## Uranium.Basement.fixed: log.radon ~ 1 + basement + uranium + (1 | county)
## Uranium.Basement.random: log.radon ~ 1 + basement + uranium + (1 + basement | county)
##          npar    AIC    BIC  logLik -2*log(L)   Chisq Df
## Uranium.Basement.fixed     5 2144.2 2168.3 -1067.1    2134.2
## Uranium.Basement.random    7 2142.6 2176.4 -1064.3    2128.6 5.5459  2
##                      Pr(>Chisq)
## Uranium.Basement.fixed
## Uranium.Basement.random    0.06248 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model comparison

Different random part (with random slope)

Model	Equation	Deviance	df
\mathcal{M}_0	$Y_{ij} = \gamma_1 + \gamma_2 \text{Uranium}_j + \beta_{1j} \text{Basement}_{ij} + U_{0j} + R_{ij}$	$D_0 = 2134.2$	$v_0 = 5$
\mathcal{M}_1	$Y_{ij} = \gamma_1 + \gamma_2 \text{Uranium}_j + \beta_{1j} \text{Basement}_{ij} + U_{0j} + U_{1j} \text{Basement}_{ij} + R_{ij}$	$D_1 = 2128.6$	$v_1 = 7$

- ▶ Test statistic: $D_0 - D_1 \sim \chi^2_2$
- ▶ $D_0 - D_1 = 2134.2 - 2128.6 = 5.5459$
- ▶ $p = P[D_0 - D_1 > 5.5459] = 0.0628$

Conclusion

At $\alpha = 5\%$, we fail to reject \mathcal{M}_0 .

There is not enough evidence that adding random slopes for Basement improves model fit.

We decide to retain \mathcal{M}_0 .

Versions and codes

#Versions and codes

Package	Version
knitr	1.50
lme4	1.1.38
performance	0.15.3
nlme	3.1.168
ggplot2	3.5.2
ggh4x	0.3.1

R version 4.5.1

All codes and dataset are available in [github](#)