STA 610L: Module 2.3

Random effects ANOVA (illustration)

Dr. Olanrewaju Michael Akande

STA 610L

Radon study

Recall: we want to estimate the distribution of radon levels in houses i within the 85 counties j.

There are 919 total observations in the data. The data is in the file `Radon.txt` on Sakai.

Variable	Description
radon	radon levels for each house
log_radon	log(radon)
state	state
floor	lowest living area of each house: 0 for basement, 1 for first floor
countyname	county names
countyID	ID for the county names (1-85)
fips	state + county fips code
uranium	county-level soil uranium
log_uranium	log(uranium)

RADON STUDY

```r
Radon <- read.csv("data/Radon.txt", header = T,sep="")
Radon$floor <- factor(Radon$floor,levels=c(0,1),labels=c("Basement","First Floor"))
str(Radon)
```

```
## 'data.frame':	919 obs. of  9 variables:
##  $ radon      : num  2.2 2.2 2.9 1 3.1 2.5 1.5 1 0.7 1.2 ...
##  $ state      : chr  "MN" "MN" "MN" "MN" ...
##  $ log_radon  : num  0.788 0.788 1.065 0 1.131 ...
##  $ floor      : Factor w/ 2 levels "Basement","First Floor": 2 1 1 1 1 1 1 1 1 1 ...
##  $ countyname : chr  "AITKIN" "AITKIN" "AITKIN" "AITKIN" ...
##  $ countyID   : int  1 1 1 1 2 2 2 2 2 2 ...
##  $ fips       : int  27001 27001 27001 27001 27003 27003 27003 27003 27003 27003 ...
##  $ uranium    : num  0.502 0.502 0.502 0.502 0.429 ...
##  $ log_uranium: num  -0.689 -0.689 -0.689 -0.689 -0.847 ...
```

STA 610L

Radon study

```
head(Radon)
```

```
##   radon state log_radon       floor countyname countyID  fips  uranium
## 1   2.2    MN 0.7884574 First Floor     AITKIN        1 27001 0.502054
## 2   2.2    MN 0.7884574    Basement     AITKIN        1 27001 0.502054
## 3   2.9    MN 1.0647107    Basement     AITKIN        1 27001 0.502054
## 4   1.0    MN 0.0000000    Basement     AITKIN        1 27001 0.502054
## 5   3.1    MN 1.1314021    Basement      ANOKA        2 27003 0.428565
## 6   2.5    MN 0.9162907    Basement      ANOKA        2 27003 0.428565
##   log_uranium
## 1  -0.6890476
## 2  -0.6890476
## 3  -0.6890476
## 4  -0.6890476
## 5  -0.8473129
## 6  -0.8473129
```

```
summary(Radon[,-c(2,7)])
```

```
##      radon           log_radon               floor       countyname
##  Min.   : 0.000   Min.   :-2.3026   Basement   :766   Length:919
##  1st Qu.: 1.900   1st Qu.: 0.6419   First Floor:153   Class :character
##  Median : 3.600   Median : 1.2809                     Mode  :character
##  Mean   : 4.768   Mean   : 1.2246
##  3rd Qu.: 6.000   3rd Qu.: 1.7918
##  Max.   :48.200   Max.   : 3.8754
##     countyID        uranium         log_uranium
##  Min.   : 1.00   Min.   :0.4140   Min.   :-0.88183
##  1st Qu.:21.00   1st Qu.:0.6221   1st Qu.:-0.47467
##  Median :44.00   Median :0.9080   Median :-0.09652
##  Mean   :43.52   Mean   :0.9339   Mean   :-0.13171
##  3rd Qu.:70.00   3rd Qu.:1.2011   3rd Qu.: 0.18324
##  Max.   :85.00   Max.   :1.6956   Max.   : 0.52802
```

Radon study

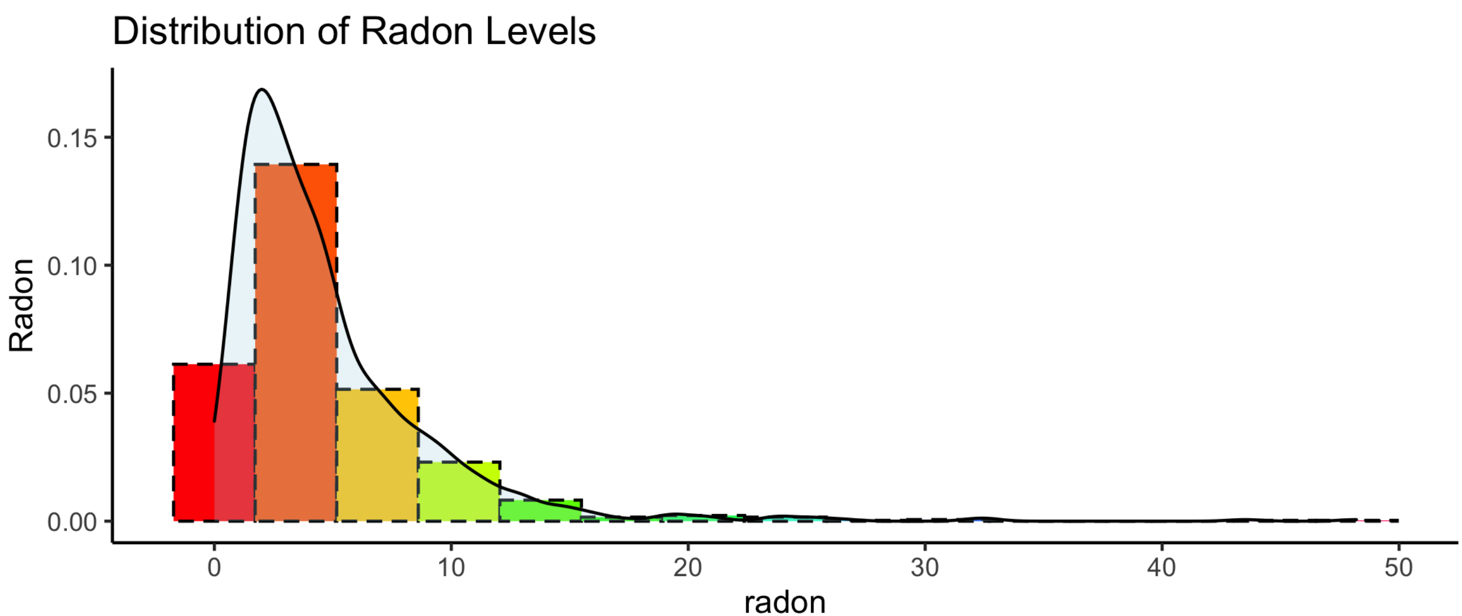
```
table(Radon$countyname) #we don't have enough data in some counties, so we should look to borrow information across counties.
```

```
## 
##         AITKIN          ANOKA         BECKER       BELTRAMI 
##              4             52              3              7 
##         BENTON      BIG STONE     BLUE EARTH          BROWN 
##              4              3             14              4 
##        CARLTON         CARVER           CASS       CHIPPEWA 
##             10              6              5              4 
##        CHISAGO           CLAY     CLEARWATER           COOK 
##              6             14              4              2 
##     COTTONWOOD      CROW WING         DAKOTA          DODGE 
##              4             12             63              3 
##        DOUGLAS      FARIBAULT       FILLMORE       FREEBORN 
##              9              6              2              9 
##        GOODHUE       HENNEPIN        HOUSTON        HUBBARD 
##             14            105              6              5 
##         ISANTI         ITASCA        JACKSON        KANABEC 
##              3             11              5              4 
##      KANDIYOHI        KITTSON    KOOCHICHING  LAC QUI PARLE 
##              4              3              7              2 
##           LAKE LAKE OF THE WOODS       LE SUEUR        LINCOLN 
##              9              4              5              4 
##           LYON       MAHNOMEN       MARSHALL         MARTIN 
##              8              1              9              7 
##         MCLEOD         MEEKER    MILLE LACS       MORRISON 
##             13              5              2              9 
##          MOWER         MURRAY       NICOLLET         NOBLES 
##             13              1              4              3 
##         NORMAN        OLMSTED     OTTER TAIL     PENNINGTON 
##              3             23              8              3 
##           PINE      PIPESTONE           POLK           POPE 
##              6              4              4              2 
##         RAMSEY        REDWOOD      RENVILLE           RICE 
##             32              5              3             11 
##           ROCK         ROSEAU          SCOTT      SHERBURNE 
##              2             14             13              8 
##         SIBLEY       ST LOUIS        STEARNS         STEELE 
##              4            116             25             10 
##        STEVENS          SWIFT           TODD       TRAVERSE 
##              2              4              3              4 
##        WABASHA         WADENA         WASECA     WASHINGTON 
##              7              5              4             46 
##       WATONWAN         WILKIN         WINONA         WRIGHT 
##              3              1             13             13 
##  YELLOW MEDICINE 
##              2
```

Radon study

The raw radon levels can only take on positive values.

```
ggplot(Radon,aes(radon)) +
  geom_histogram(aes(y=..density..),color="black",linetype="dashed",
               fill=rainbow(15),bins=15) + theme(legend.position="none") +
  geom_density(alpha=.25, fill="lightblue") + scale_fill_brewer(palette="Blues") +
  labs(title="Distribution of Radon Levels",y="Radon") + theme_classic()
```

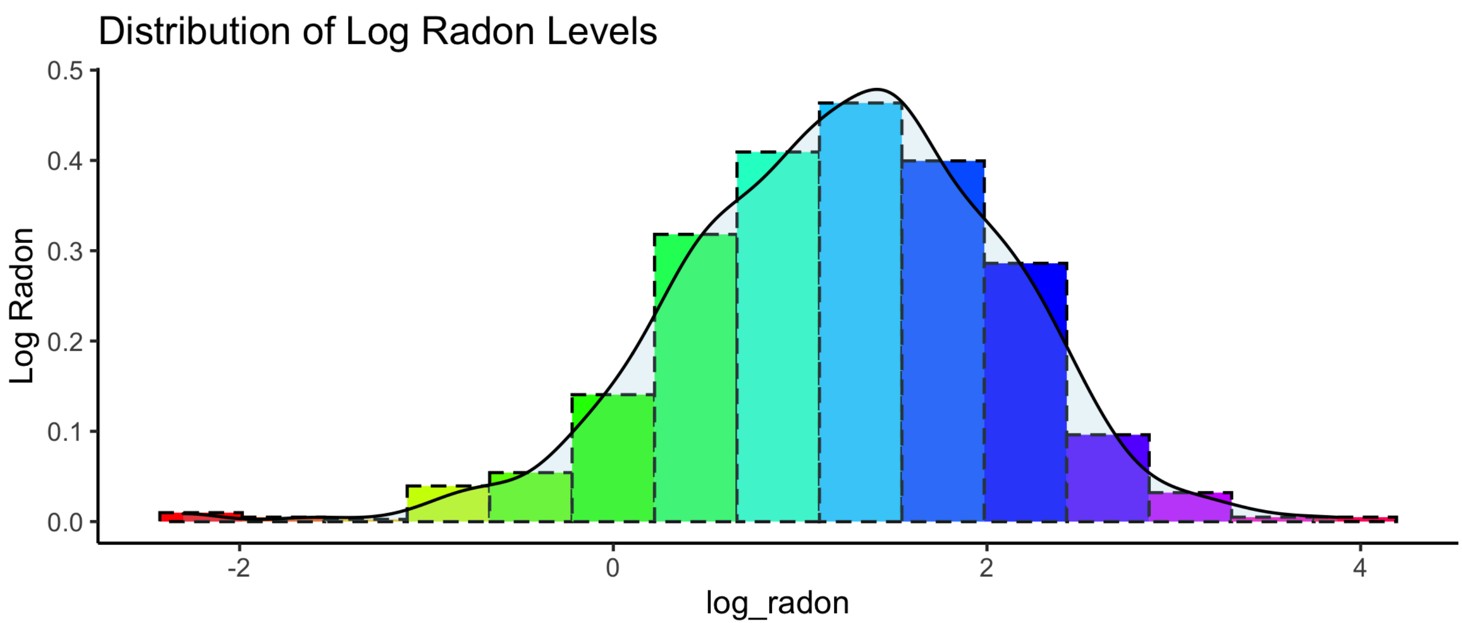


Obviously very skewed.

Radon study

Let's look at `log_radon` instead.

```
ggplot(Radon,aes(log_radon)) +
  geom_histogram(aes(y=..density..),color="black",linetype="dashed",
                 fill=rainbow(15),bins=15) + theme(legend.position="none") +
  geom_density(alpha=.25, fill="lightblue") + scale_fill_brewer(palette="Blues") +
  labs(title="Distribution of Log Radon Levels",y="Log Radon") + theme_classic()
```



Much better! Let's go with log radon for now.

STA 610L

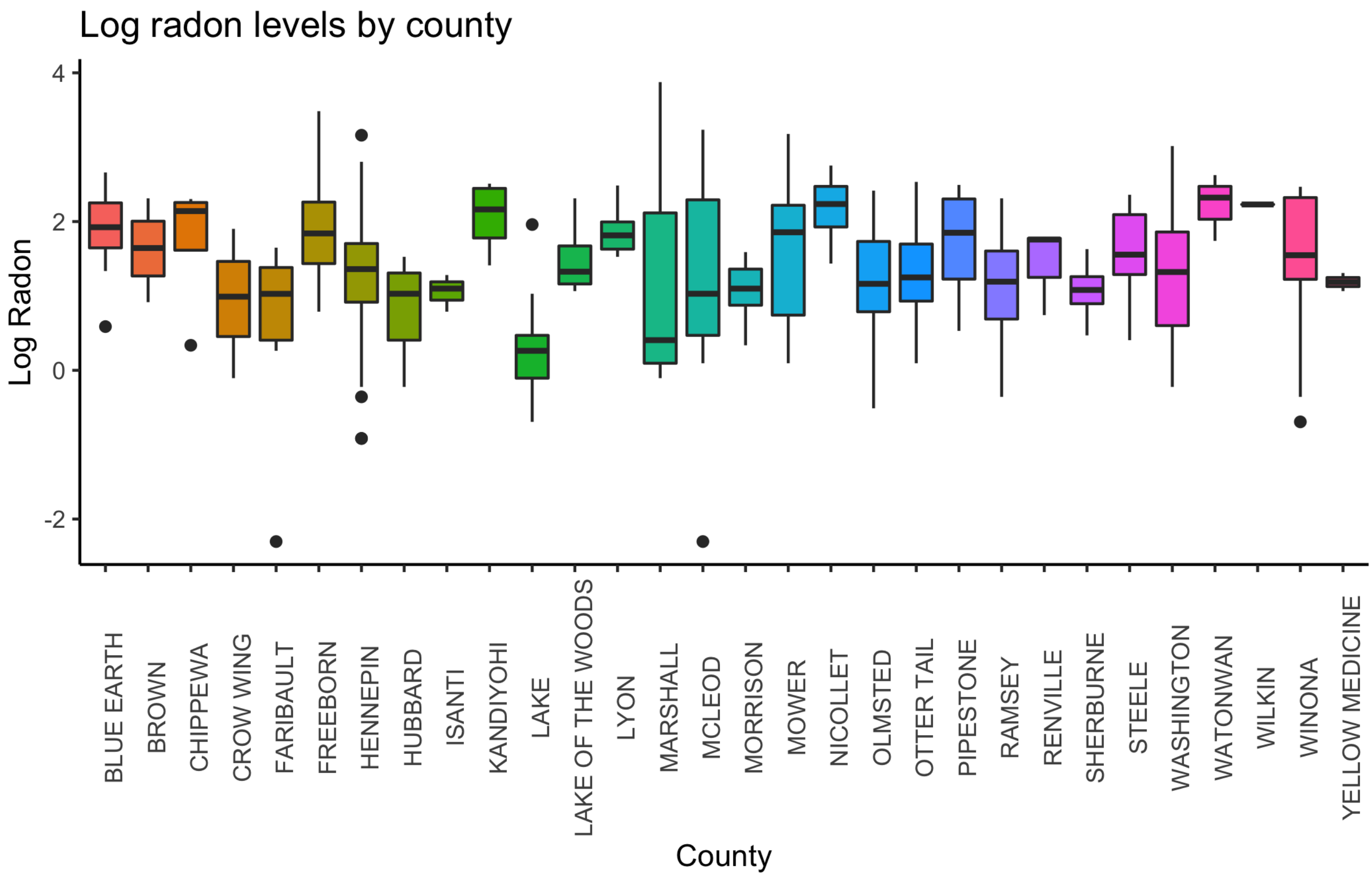
Radon study

Are there any variations of radon levels by county? There are too many counties, so, let's do it for a random sample of counties.

```
set.seed(1000)
sample_county <- sample(unique(Radon$countyname),25,replace=F)
ggplot(Radon[is.element(Radon$countyname,sample_county),],
       aes(x=countyname, y=log_radon, fill=countyname)) +
  geom_boxplot() +
  labs(title="Log radon levels by county",
       x="County",y="Log Radon") + theme_classic() +
  theme(legend.position="none",axis.text.x = element_text(angle = 90))
```

STA 610L

Radon study



Looks like the levels vary by county. However, there are many counties with very little data.

STA 610L

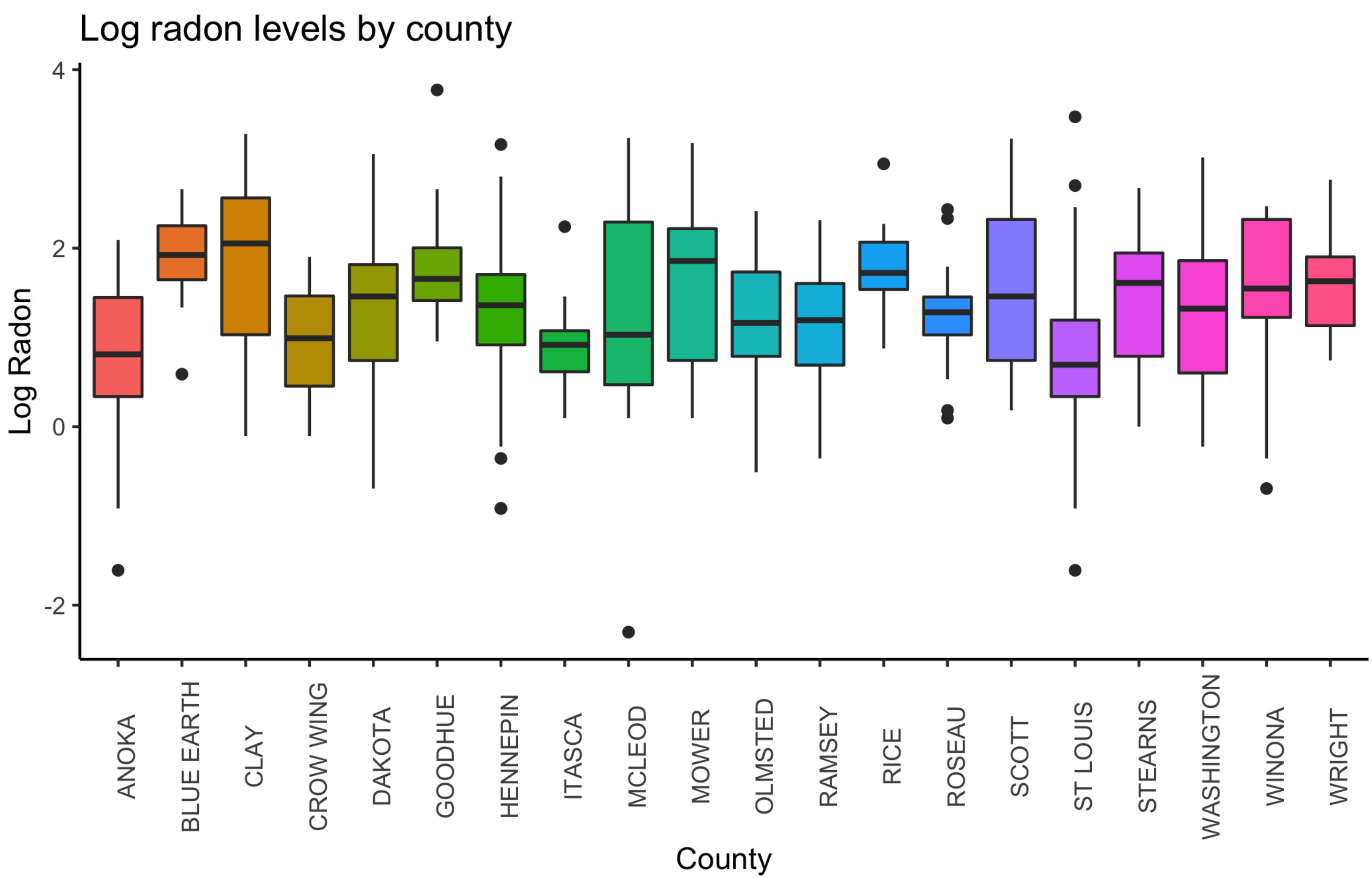
Radon study

Let's focus on counties with at least 11 houses.

```
sample_county <- which(table(Radon$countyID) > 10)
ggplot(Radon[is.element(Radon$countyID,sample_county),],
       aes(x=countyname, y=log_radon, fill=countyname)) +
  geom_boxplot() +
  labs(title="Log radon levels by county",
       x="County",y="Log Radon") + theme_classic() +
  theme(legend.position="none",axis.text.x = element_text(angle = 90))
```

STA 610L

Radon study



What can you conclude from this plot?

STA 610L

Radon study in R

We start with a random effects ANOVA model. We can fit that model in R by doing

```
Model1 <- lmer(log_radon ~ (1 | countyname), data = Radon)
summary(Model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log_radon ~ (1 | countyname)
##    Data: Radon
##
## REML criterion at convergence: 2259.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4661 -0.5734  0.0441  0.6432  3.3516
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  countyname (Intercept) 0.09581  0.3095
##  Residual               0.63662  0.7979
## Number of obs: 919, groups:  countyname, 85
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.31258    0.04891   26.84
```

STA 610L

Radon study in R

```
coef(Model1)
```

```
## $countyname
##                   (Intercept)
## AITKIN              1.0674994
## ANOKA               0.8875568
## BECKER              1.2303812
## BELTRAMI            1.2245444
## BENTON              1.2899760
## BIG STONE           1.3749235
## BLUE EARTH          1.7171954
## BROWN               1.4315991
## CARLTON             1.0833131
## CARVER              1.2608819
## CASS                1.3506019
## CHIPPEWA            1.4695309
## CHISAGO             1.1826263
## CLAY                1.6312662
## CLEARWATER          1.1867346
## COOK                1.1627174
## COTTONWOOD          1.0953027
## CROW WING           1.0736856
## DAKOTA              1.2944691
## DODGE               1.4642936
## DOUGLAS             1.5089532
## FARIBAULT           0.9349171
## FILLMORE            1.2494467
## FREEBORN            1.6743898
## GOODHUE             1.6755316
## HENNEPIN            1.2867314
## HOUSTON             1.4172565
## HUBBARD             1.0964581
## ISANTI              1.2327649
## ITASCA              1.0714227
## JACKSON             1.6165831
## KANABEC             1.2839088
## KANDIYOHI           1.5941519
## KITTSON             1.2495867
## KOOCHICHING         0.8482223
## LAC QUI PARLE       1.6101391
## LAKE                0.7427261
## LAKE OF THE WOODS   1.3858053
## LE SUEUR            1.4376932
## LINCOLN             1.6218834
## LYON                1.6209590
## MAHNOMEN            1.3189074
## MARSHALL            1.2489275
## MARTIN              1.1204333
## MCLEOD              1.1545020
## MEEKER              1.2705104
## MILLE LACS          1.1300481
## MORRISON            1.1719088
## MOWER               1.4969979
## MURRAY              1.4670206
## NICOLLET            1.6329198
## NOBLES              1.5039156
```

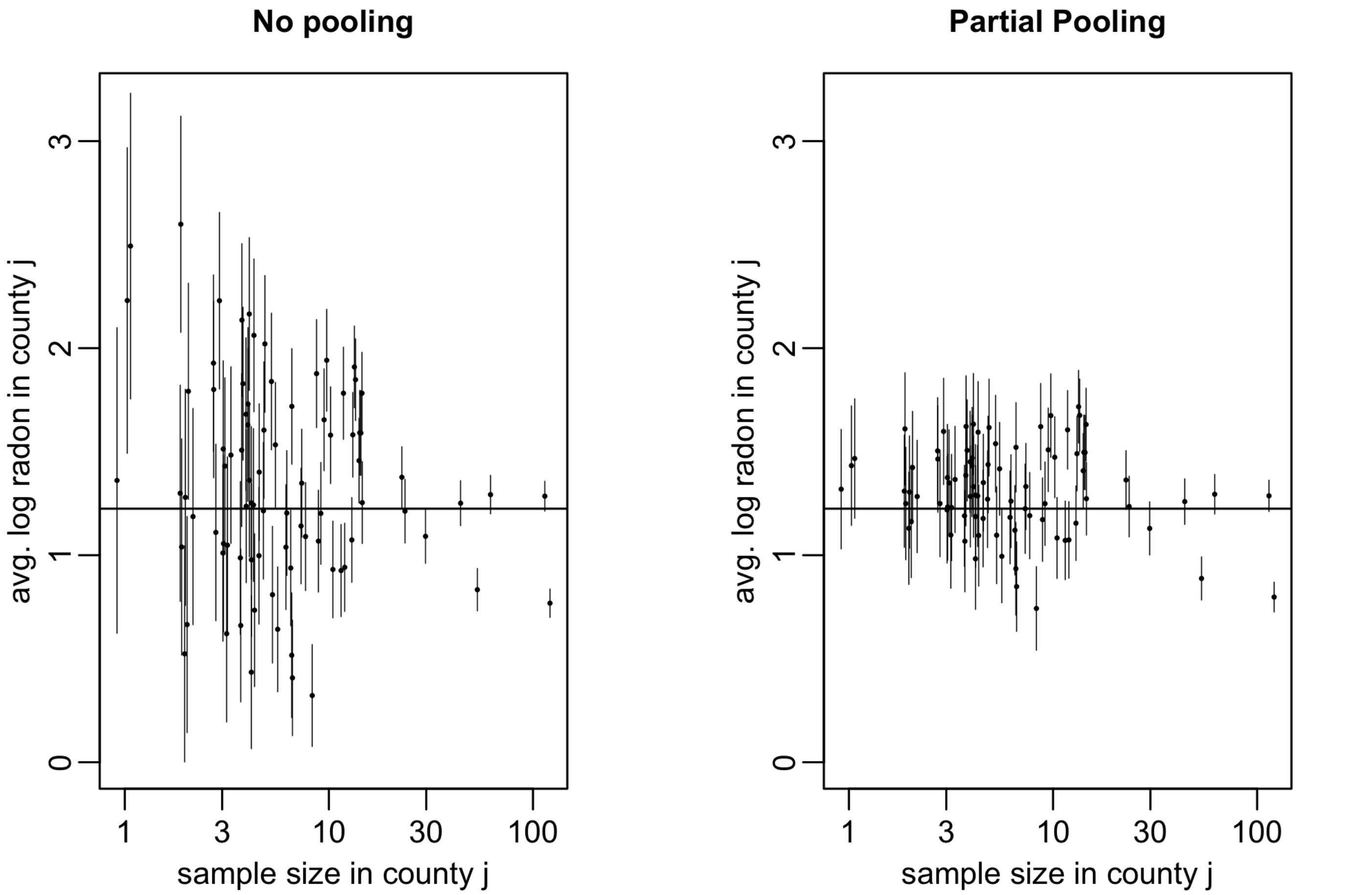
STA 610L

Plotting code

```
y <- Radon$log_radon; ybarbar <- mean(y)
J <- length(unique(Radon$countyname))
sample_size <- as.vector(table(Radon$countyname))
sample_size_jittered <- sample_size*exp(runif(J, -.1, .1))

cty_mns <- tapply(y,Radon$countyname,mean)
cty_vars <- tapply(y,Radon$countyname,var)
cty_sds <- mean(sqrt(cty_vars[!is.na(cty_vars)]))/sqrt(sample_size)
cty_sds_sep <- sqrt(tapply(y,Radon$countyname,var)/sample_size)


par(mfrow=c(1,2))
plot (sample_size_jittered, cty_mns, cex.lab=.9, cex.axis=1,
      xlab="sample size in county j",
      ylab="avg. log radon in county j",
      pch=20, log="x", cex=.3, mgp=c(1.5,.5,0),
      ylim=c(0,3.2), yaxt="n", xaxt="n")
axis (1, c(1,3,10,30,100), cex.axis=.9, mgp=c(1.5,.5,0))
axis (2, seq(0,3), cex.axis=.9, mgp=c(1.5,.5,0))
for (j in 1:J){
  lines (rep(sample_size_jittered[j],2),
         cty_mns[j] + c(-1,1)*cty_sds[j], lwd=.5)
#         cty_mns[j] + c(-1,1)*mean(cty_sds[!is.na(cty_sds)]), lwd=.5)
}
title("No pooling",cex.main=.9, line=1)
abline(h=ybarbar)


library(arm)
plot (sample_size_jittered, coef(Model1)$countyname[,1], cex.lab=.9, cex.axis=1,
      xlab="sample size in county j",
      ylab="avg. log radon in county j",
      pch=20, log="x", cex=.3, mgp=c(1.5,.5,0),
      ylim=c(0,3.2), yaxt="n", xaxt="n")
axis (1, c(1,3,10,30,100), cex.axis=.9, mgp=c(1.5,.5,0))
axis (2, seq(0,3), cex.axis=.9, mgp=c(1.5,.5,0))
for (j in 1:J){
  lines (rep(sample_size_jittered[j],2),coef(Model1)$countyname[j,1] + c(-1,1)*se.ranef(Model1)$countyname[j,1], lwd=.5)
}
title("Partial Pooling",cex.main=.9, line=1)
abline(h=ybarbar)
```

STA 610L

Plots



No pooling
Partial Pooling
avg. log radon in county j
sample size in county j
0
1
2
3
1
3
10
30
100

Random effects model plus a grouping factor

One important predictor of radon levels is the floor on which the measurement is taken: basement or first floor.

Radon comes from underground and can enter more easily when the house is built into the ground.

In addition, basements tend to have higher levels than ground floors.

So, let's explore the relationship between `log_radon` and `floor`.

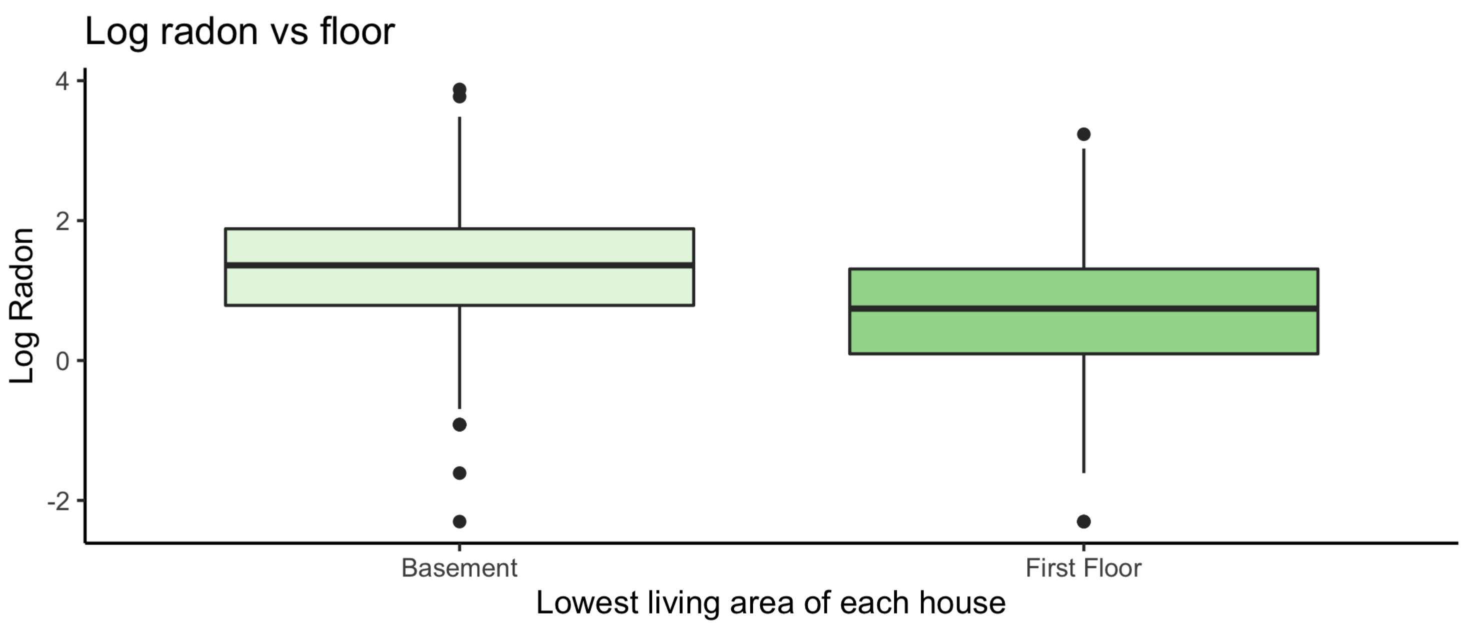
Note that is an individual-level (different observation for each house) variable, unlike the `uranium` variable, which is county-level (group-level).

We will return to this point later.

STA 610L

Radon study

```
ggplot(Radon,aes(x=floor, y=log_radon, fill=floor)) +
  geom_boxplot() + scale_fill_brewer(palette="Greens") +
  labs(title="Log radon vs floor", x="Lowest living area of each house",y="Log Radon") +
  theme_classic() + theme(legend.position="none")
```



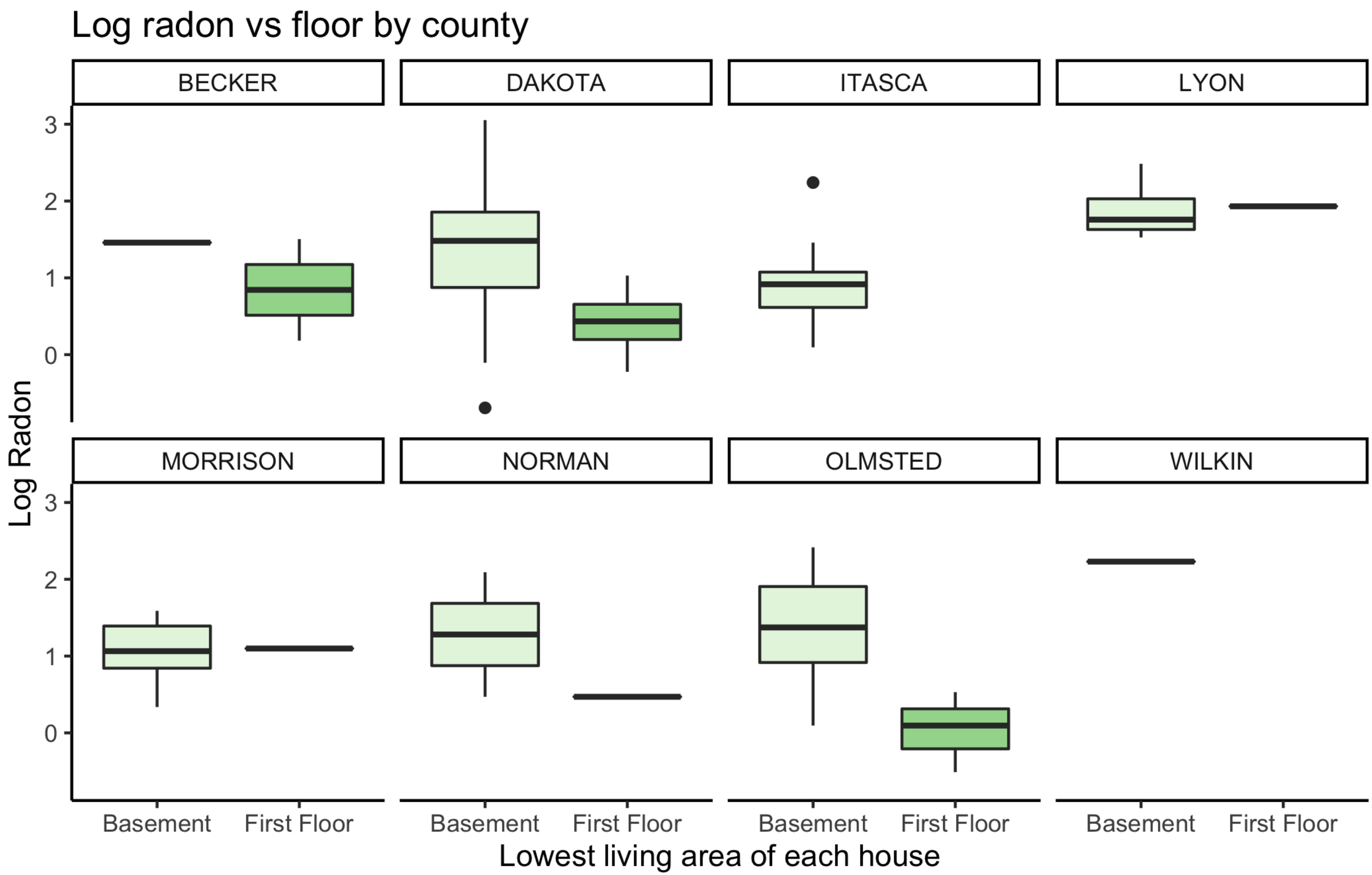
Looks like radon levels are indeed higher for houses with the basement as the lowest living area.

Radon study

Let's look at the same relationship for a random sample of counties.

```
sample_county <- sample(unique(Radon$countyname),8,replace=F)
ggplot(Radon[is.element(Radon$countyname,sample_county),],
       aes(x=floor, y=log_radon, fill=floor)) +
  geom_boxplot() +
  scale_fill_brewer(palette="Greens") +
  labs(title="Log radon vs floor by county",
       x="Lowest living area of each house",y="Log Radon") +
  theme_classic() + theme(legend.position="none") +
  facet_wrap( ~ countyname,ncol=4)
```

Radon study



Again, not enough data for some counties.

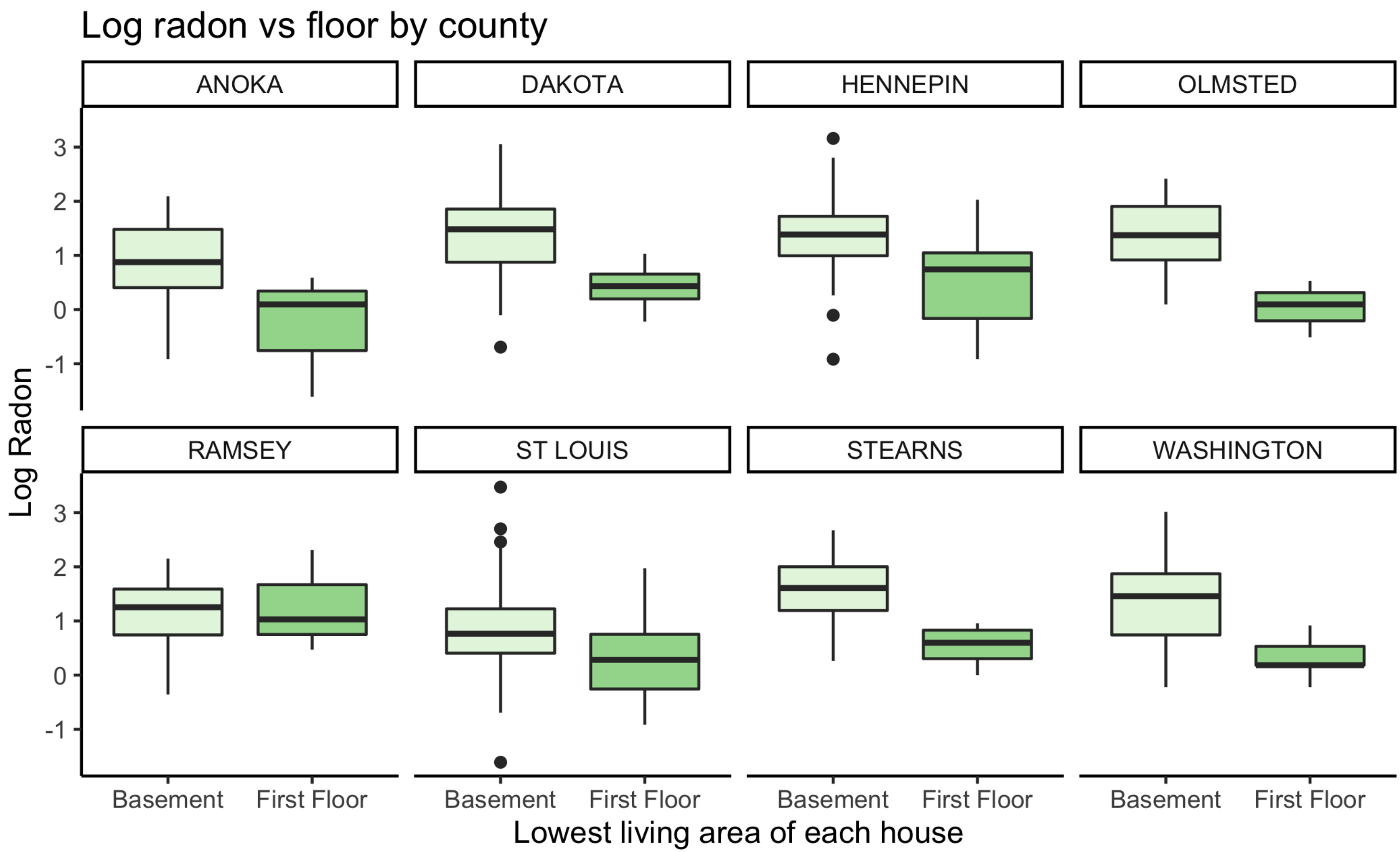
Radon study

Let's focus on counties with at least 16 houses.

```
sample_county <- which(table(Radon$countyID) > 15)
ggplot(Radon[is.element(Radon$countyID,sample_county),],
       aes(x=floor, y=log_radon, fill=floor)) +
  geom_boxplot() +
  scale_fill_brewer(palette="Greens") +
  labs(title="Log radon vs floor by county",
       x="Lowest living area of each house",y="Log Radon") +
  theme_classic() + theme(legend.position="none") +
  facet_wrap( ~ countyname,ncol=4)
```

STA 610L

Radon study



Even though the overall direction is the same, it looks like the actual differences between floor = 0 and floor = 1 differs for some counties.

STA 610L

Radon study

As we have been doing, let's begin by examining the complete-pooling regression,

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

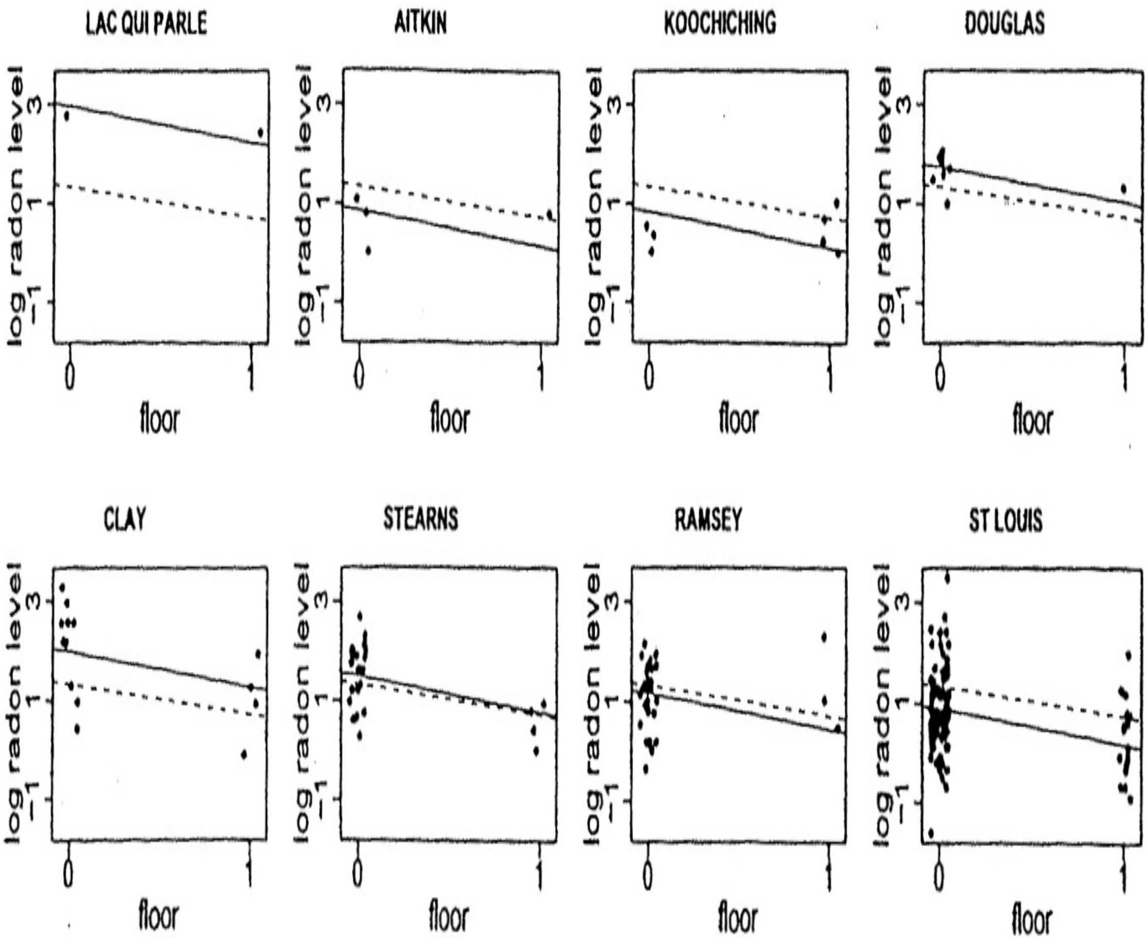
and the no-pooling regression

$$y_{ij} = \alpha_j + \beta x_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

where α_j is the mean log radon level from basement measures of homes (indexed by i) in county j.

The following plot shows the dashed lines $\hat{y} = \widehat{\alpha} + \widehat{\beta} x$ for eight selected counties from the complete pooling model, and the solid lines $\hat{y} = \widehat{\alpha}_j + \widehat{\beta} x$ from no pooling model.

Radon study



LAC QUI PARLE
AITKIN
KOOCHICHING
DOUGLAS
CLAY
STEARNS
RAMSEY
ST LOUIS
log radon level
floor

Radon study

The estimates of β (the association between floor of home and radon level) differ slightly for the two regressions, with $\widehat{\beta} = -0.61$ for the pooling model, and $\widehat{\beta} = -0.72$ for the no-pooling model.

As we might expect, we tend to have higher radon levels in the basement $(p < 0.0001)$.

As expected, neither analysis is perfect.

The complete-pooling analysis ignores variation in radon levels between counties, which is undesirable because our goal is to identify counties with high-radon homes -- we can't pool away the main research question!

The no-pooling analysis is also problematic -- for example the Lac Qui Parle County line is estimated based on just two data points.

Radon study

So we turn to a simple multilevel model instead:

$$y_{ij} = \gamma_0 + \alpha_j + \beta x_{ij} + \varepsilon_{ij},$$

where now $\alpha_j \sim N(0, \tau^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

This model can also be parameterized as

$$y_{ij} \sim N(\alpha_j + \beta x_{ij}, \sigma^2),$$

with each

$$\alpha_j \sim N\left(\gamma_0, \tau^2\right).$$

STA 610L

Radon study in R

First, the pooled model

```
###pooled model
lm_pooled <- lm(log_radon ~ floor, data = Radon)
summary(lm_pooled)
```

```
##
## Call:
## lm(formula = log_radon ~ floor, data = Radon)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6293 -0.5383  0.0342  0.5603  2.5486
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.32674    0.02972  44.640   <2e-16
## floorFirst Floor -0.61339    0.07284  -8.421   <2e-16
##
## Residual standard error: 0.8226 on 917 degrees of freedom
## Multiple R-squared:  0.07178,    Adjusted R-squared:  0.07077
## F-statistic: 70.91 on 1 and 917 DF,  p-value: < 2.2e-16
```

STA 610L

Radon study in R

Next, the unpooled model.

```
###unpooled  model
lm_unpooled <- lm(log_radon ~ floor + countyname, data = Radon)
summary(lm_unpooled)
```

```
##
## Call:
## lm(formula = log_radon ~ floor + countyname, data = Radon)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.14595 -0.45405  0.00065  0.45376  2.65987
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.84054    0.37866   2.220  0.02670
## floorFirst Floor            -0.72054    0.07352  -9.800  < 2e-16
## countynameANOKA              0.03428    0.39274   0.087  0.93047
## countynameBECKER             0.68816    0.57854   1.189  0.23459
## countynameBELTRAMI           0.71218    0.47470   1.500  0.13392
## countynameBENTON             0.59203    0.53487   1.107  0.26867
## countynameBIG STONE          0.67247    0.57802   1.163  0.24500
## countynameBLUE EARTH         1.17162    0.42892   2.732  0.00644
## countynameBROWN              1.14904    0.53519   2.147  0.03208
## countynameCARLTON            0.16250    0.44764   0.363  0.71669
## countynameCARVER             0.72336    0.48861   1.480  0.13913
## countynameCASS               0.56059    0.50775   1.104  0.26988
## countynameCHIPPEWA           0.88971    0.53519   1.662  0.09680
## countynameCHISAGO            0.19818    0.48861   0.406  0.68514
## countynameCLAY               1.14784    0.42886   2.677  0.00759
## countynameCLEARWATER         0.49743    0.53519   0.929  0.35292
```

STA 610L

Radon study in R

Finally, the random intercepts model with a fixed effect for floor.

```
Model2 <- lmer(log_radon ~ floor + (1 | countyname), data = Radon)
summary(Model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log_radon ~ floor + (1 | countyname)
##    Data: Radon
##
## REML criterion at convergence: 2171.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3989 -0.6155  0.0029  0.6405  3.4281
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  countyname (Intercept) 0.1077   0.3282
##  Residual               0.5709   0.7556
## Number of obs: 919, groups:  countyname, 85
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       1.46160    0.05158  28.339
## floorFirst Floor -0.69299    0.07043  -9.839
##
## Correlation of Fixed Effects:
##             (Intr)
## florFrstFlr -0.288
```

STA 610L

Radon study in R

```
coef(Model2)
```

```
## $countyname
##                   (Intercept) floorFirst Floor
## AITKIN              1.1915003       -0.6929937
## ANOKA               0.9276468       -0.6929937
## BECKER              1.4792143       -0.6929937
## BELTRAMI            1.5045012       -0.6929937
## BENTON              1.4461503       -0.6929937
## BIG STONE           1.4801817       -0.6929937
## BLUE EARTH          1.8581255       -0.6929937
## BROWN               1.6827736       -0.6929937
## CARLTON             1.1600746       -0.6929937
## CARVER              1.5086099       -0.6929937
## CASS                1.4322449       -0.6929937
## CHIPPEWA            1.5771520       -0.6929937
## CHISAGO             1.2370518       -0.6929937
## CLAY                1.8380232       -0.6929937
## CLEARWATER          1.4024982       -0.6929937
## COOK                1.2432992       -0.6929937
## COTTONWOOD          1.3723633       -0.6929937
## CROW WING           1.2209415       -0.6929937
## DAKOTA              1.3462611       -0.6929937
## DODGE               1.5840333       -0.6929937
## DOUGLAS             1.6311136       -0.6929937
## FARIBAULT           1.0211902       -0.6929937
## FILLMORE            1.4409443       -0.6929937
## FREEBORN            1.8605721       -0.6929937
## GOODHUE             1.8135585       -0.6929937
## HENNEPIN            1.3626875       -0.6929937
## HOUSTON             1.6222663       -0.6929937
## HUBBARD             1.3467692       -0.6929937
## ISANTI              1.3149878       -0.6929937
## ITASCA              1.0999775       -0.6929937
## JACKSON             1.7329563       -0.6929937
## KANABEC             1.3646863       -0.6929937
## KANDIYOHI           1.7197951       -0.6929937
## KITTSON             1.5015319       -0.6929937
## KOOCHICHING         1.0870316       -0.6929937
## LAC QUI PARLE       1.8680900       -0.6929937
## LAKE                0.7928241       -0.6929937
## LAKE OF THE WOODS   1.6303574       -0.6929937
## LE SUEUR            1.5979923       -0.6929937
## LINCOLN             1.8260565       -0.6929937
## LYON                1.7636308       -0.6929937
## MAHNOMEN            1.4456250       -0.6929937
## MARSHALL            1.5404841       -0.6929937
## MARTIN              1.2199767       -0.6929937
## MCLEOD              1.3375197       -0.6929937
## MEEKER              1.3416955       -0.6929937
## MILLE LACS          1.2995480       -0.6929937
## MORRISON            1.2623707       -0.6929937
## MOWER               1.6294468       -0.6929937
## MURRAY              1.6253581       -0.6929937
## NICOLLET            1.7641694       -0.6929937
## NOBLES              1.6300755       -0.6929937
```

Interpretation of fixed effects

- Intuitively, we have an overall "average" regression line for all houses across all counties in Minnesota which has slope -0.69 and intercept 1.46.
- That is, the general estimated line for any of the houses in Minnesota is:

$$\widehat{\log(\text{radon})}_{ij} = 1.46 - 0.69 \times \text{floor}_{ij}$$

- For any house in Minnesota with a basement as the lowest living area, the baseline radon level is $e^{1.46} = 4.31$.
- Then, for any house in Minnesota, having a first floor as the lowest living area, instead of a basement, reduces the radon level by a multiplicative effect of $e^{-0.69} \approx 0.5016$, that is, about a 49.84% reduction.
- However, if the house is in Dakota county for example, we also need to add on the random intercept for that county.

STA 610L

Interpretation of fixed effects

- For Dakota county, we have

```
(ranef(Model2)$countyname)["DAKOTA",]
```

```
## [1] -0.1153368
```

so that the estimated regression line for Dakota county is approximately

$$\widehat{\log(\text{radon})}_{ij} = (1.46 - 0.12) + 0.69 \times \text{floor}_{ij} = 1.35 - 0.69 \times \text{floor}_{ij}$$

- Thus, for any house in Dakota county in Minnesota with a basement as the lowest living area, the baseline radon level is actually $e^{1.34} \approx 3.82$, which is lower than the overall state wide average.

- For floor effect remains the same, at least until we explore random slopes.

STA 610L

Radon study in R

Again,

```
summary(Model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log_radon ~ floor + (1 | countyname)
##    Data: Radon
##
## REML criterion at convergence: 2171.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3989 -0.6155  0.0029  0.6405  3.4281
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  countyname (Intercept) 0.1077   0.3282
##  Residual               0.5709   0.7556
## Number of obs: 919, groups:  countyname, 85
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)       1.46160    0.05158  28.339
## floorFirst Floor -0.69299    0.07043  -9.839
##
## Correlation of Fixed Effects:
##             (Intr)
## florFrstFlr -0.288
```

STA 610L

Interpretation of random effects

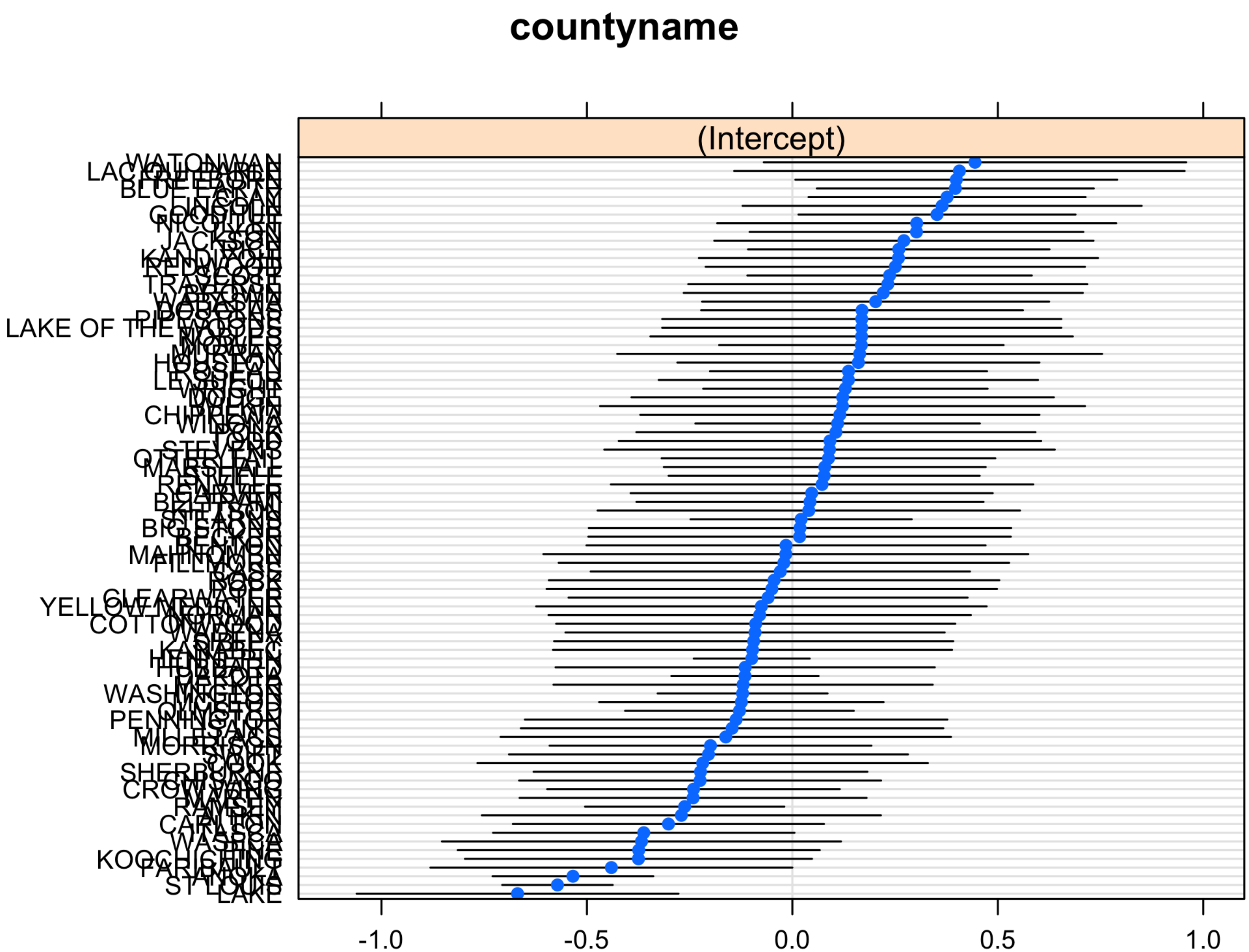
- The estimated standard error $\hat{\sigma} = 0.76$ describes the within-county or remaining unexplained variation.
- The estimated $\hat{\tau} = 0.33$ describes the across-county variation attributed to the random intercept.
- The correlation between two houses in the same county is then given by

$$\widehat{\text{Corr}(y_{ij}, y_{i'j})} = \frac{\hat{\tau}^2}{\hat{\sigma}^2 + \hat{\tau}^2} = \frac{0.1077}{0.5709 + 0.1077} \approx 0.16.$$

- We do have some correlation, but not that strong.
- You can visualize the random effects by typing `dotplot(ranef(Model2, condVar=TRUE))$countyname` in R.
- So many counties! So, you will need to zoom out on your computer.

STA 610L

Interpretation of random effects



What's next?

Move on to the readings for the next module!