

Table 1: Estimated Coefficients from Logit Regression of Water Well Switching.

	Model 1	Model 2
(Intercept)	0.85 (0.53)	1.40 (0.92)
arsenic	-1.22 (0.51)*	-1.77 (0.91)
dist100	-0.46 (0.10)***	-0.97 (0.71)
arsenic:dist100		0.51 (0.70)
Deviance	1167.76	1167.23
Num. obs.	1000	1000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Question 1

Many of the wells used for drinking water in Bangladesh and other South Asian countries are contaminated with natural arsenic, affecting an estimated 100 million people. Arsenic is a cumulative poison, and exposure increases the risk of cancer and other diseases, with risks estimated to be proportional to exposure.

A research team measured all the wells and labeled them with their arsenic level as well as a characterization as “safe” (below 0.5 in units of hundreds of micrograms per liter, the Bangladesh standard for arsenic in drinking water) or “unsafe” (above 0.5). People with unsafe wells were encouraged to switch to nearby private or community wells or to new wells of their own construction. A few years later, the researchers returned to find out who had switched wells.

We performed a logistic regression analysis with the data to understand the factors that predict well switching among the users of unsafe wells. Your outcome variable *switch* is 1 if household i switched to a new well, and 0 if household i continued using its own well.

We estimated models with the following inputs:

- The distance (in meters/100) to the closest known safe well
 - The arsenic level of respondent’s well
- (a) So, we successfully estimated an additive model with arsenic and distance to the nearest safe well as the two predictors of whether a household switched to a new well. The estimated coefficients are found below in Table 1. Interpret the estimated coefficients for the intercept and each predictor.
 - (b) Should we estimate a different effect for the distance to the closest known safe well on switching wells when there is arsenic in one’s own well? If so, change your interpretation of the estimated coefficients in part (a) to conform with the interactive model in Table 1. Provide the appropriate test to determine whether we should model the relationship between distance, arsenic, and switching wells using an additive or interactive model.
 - (c) Compute the average difference in the probability of switching wells between two households that have arsenic = 0.75, but one is closer (dist100 = 0.37) than the other (dist100 = 2.07).

Question 3

Radon is a naturally occurring carcinogen, and the distribution of radon levels in U.S. homes varies greatly, with some houses that have dangerously high concentrations. For the purpose of this analysis, the data were structured hierarchically: houses within counties. The radon data in Figure 1 shows the logarithm of the home radon measurement versus floor of measurement for houses sampled from eight of the 85 counties in Minnesota. In each graph of Figure 1, the dashed line shows the linear regression of log radon, given the floor of measurement, using a model that pools all counties together (so the same line appears in all eight plots), and the solid line shows the no-pooling regressions, obtained by including county indicators in the regression.

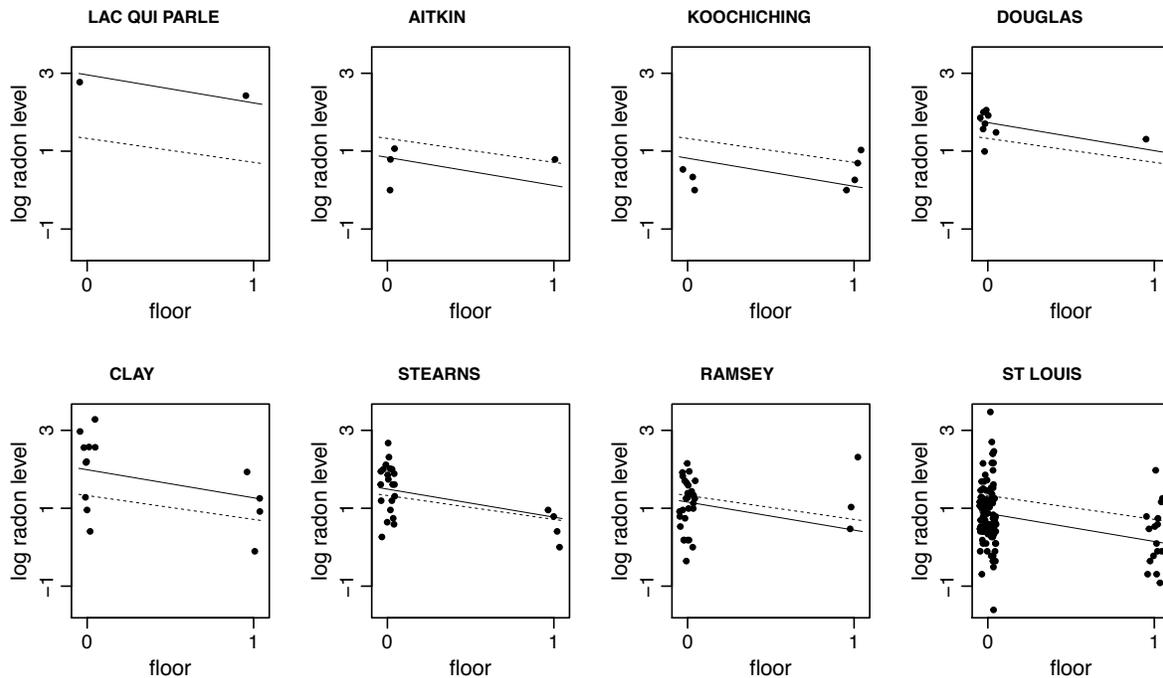


Figure 1: Radon levels in eight Minnesota counties.

- a) Now, let's say we fit a partially pooled model in which we allow for varying intercepts. Interpret the R output below to determine whether we should estimate the across county variation in intercepts, whether it's sufficient to model the within county variation in intercepts, or whether we do not need to model within or across county variation.

```
lmer(formula = y ~ x + (1 | county))
      coef.est coef.se
(Intercept)  1.46    0.05
x            -0.69    0.07
Error terms:
Groups   Name      Std.Dev.
county  (Intercept) 0.33
Residual                    0.76
# of obs: 919, groups: county, 85
deviance = 2163.7
```

- b) Describe how the fitted regression lines for the partial pooling model would be different in comparison to the completely pooled or un-pooled models.