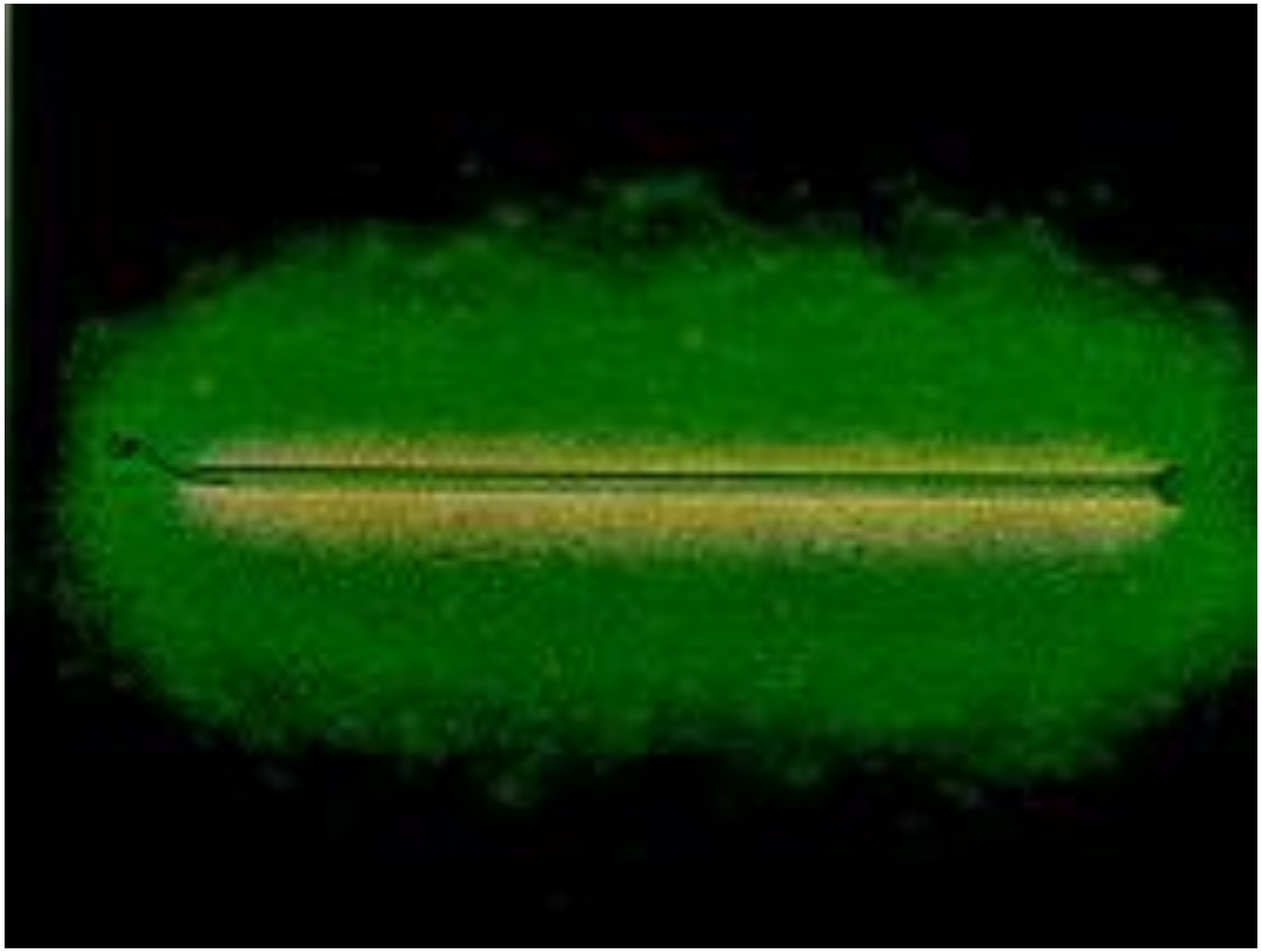
Taking advantage of within-group and between-group information in linear regressions: Multilevel modeling and partial pooling

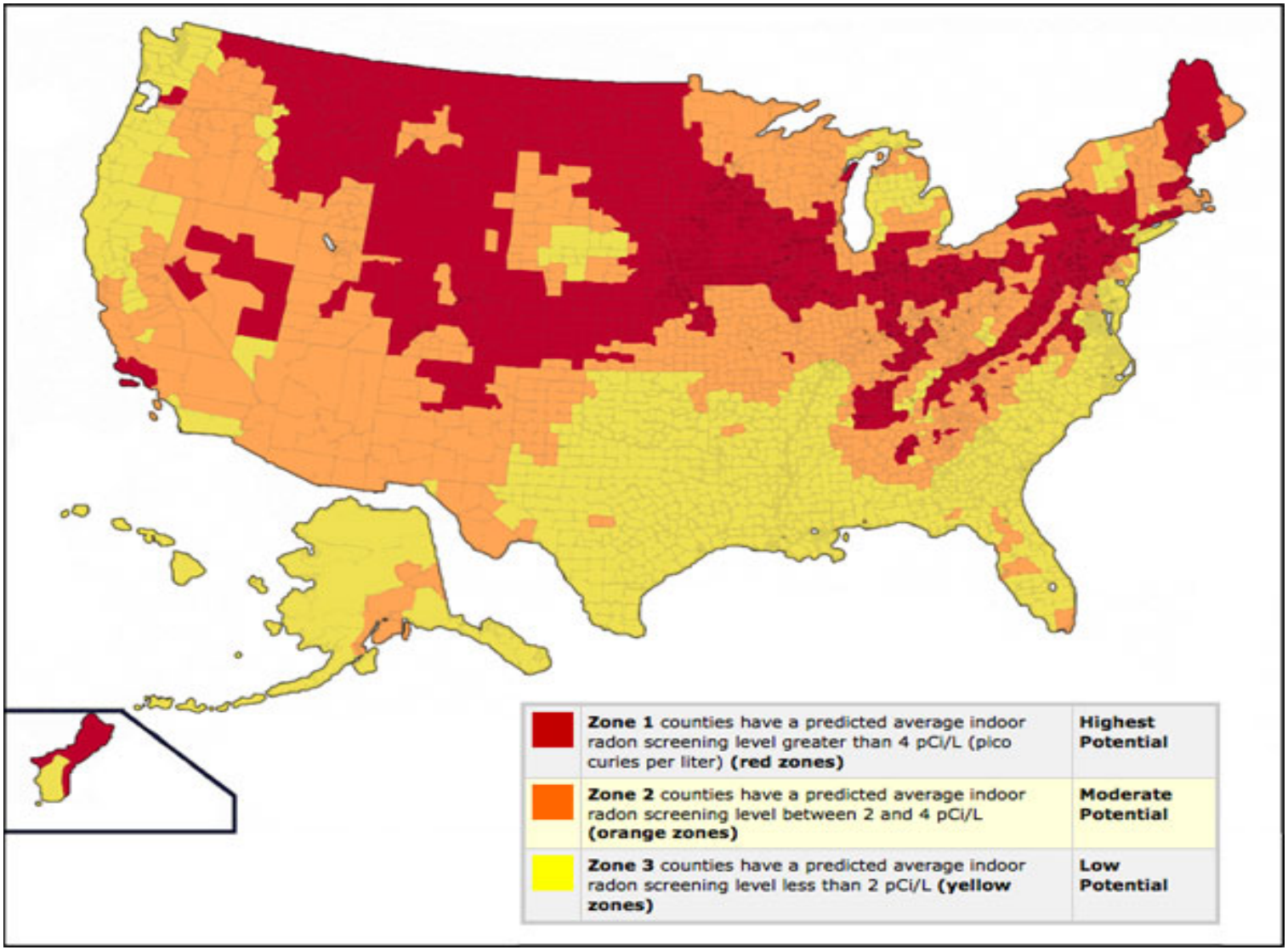
Michael Maltese

27 Sep. 2013

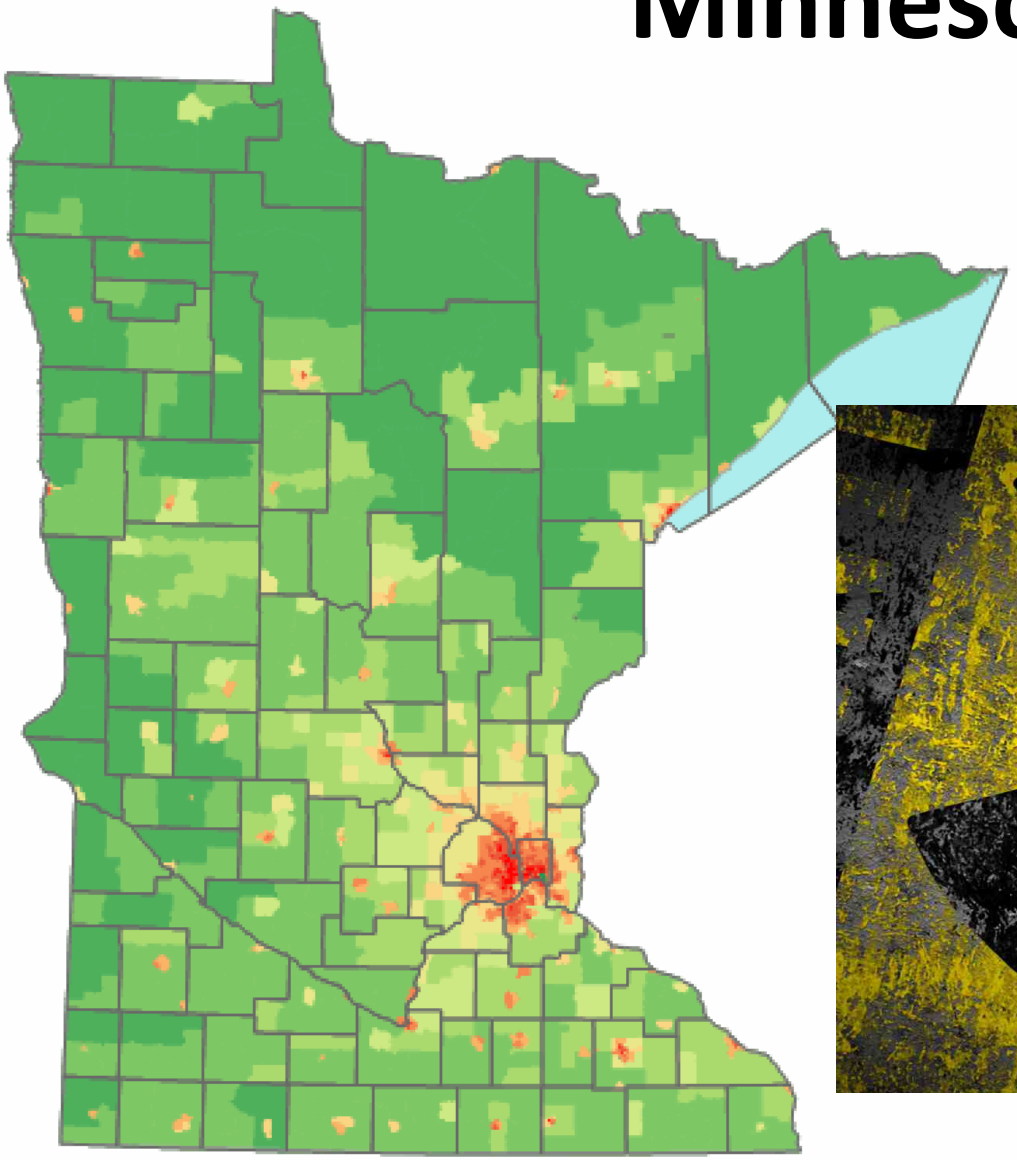
Let’s talk about radon!



Where is it??


Zone 1 counties have a predicted average indoor radon screening level greater than 4 pCi/L (pico curies per liter) (red zones)
Highest Potential
Zone 2 counties have a predicted average indoor radon screening level between 2 and 4 pCi/L (orange zones)
Moderate Potential
Zone 3 counties have a predicted average indoor radon screening level less than 2 pCi/L (yellow zones)
Low Potential

Minnesota



Data: The HIGH-RADON Project

- From *Lawrence Berkeley National Laboratory*
- Radon measurements in houses for every county in the contiguous 48 states

IMPORTANT: Do you have a basement? Radon often comes in through your basement. If you have a basement get rid of it.

Where do you not want to live?

What is a regression?

$$Y = \beta X + \epsilon$$

Idea: some data is function of other data. To take an example from economics, wage might be a function of how old you are and how much education you have. Seems sort of reasonable, right?

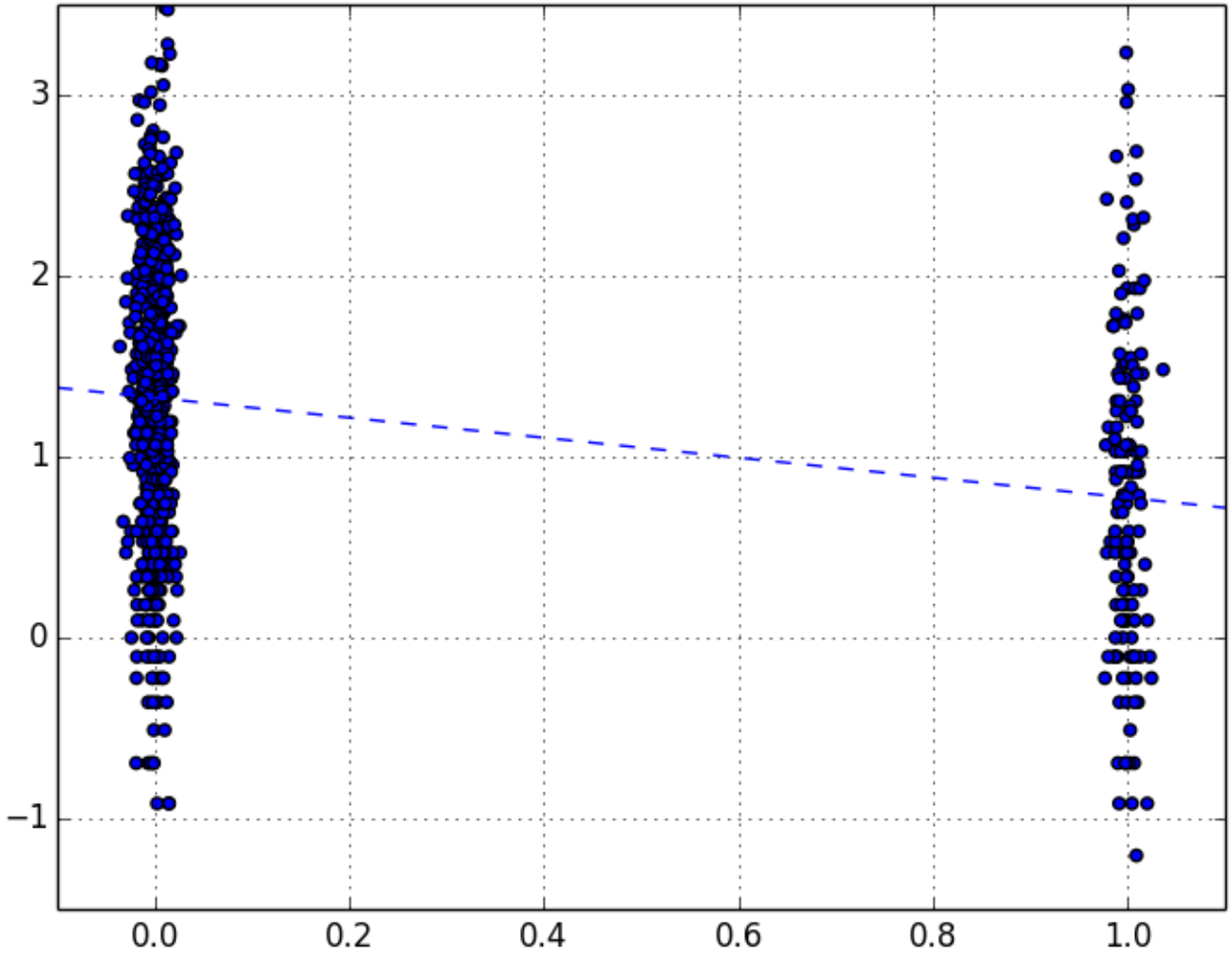
Pooled: all of Minnesota

$$radon = \beta_0 + \beta_1 basement?? + \epsilon$$

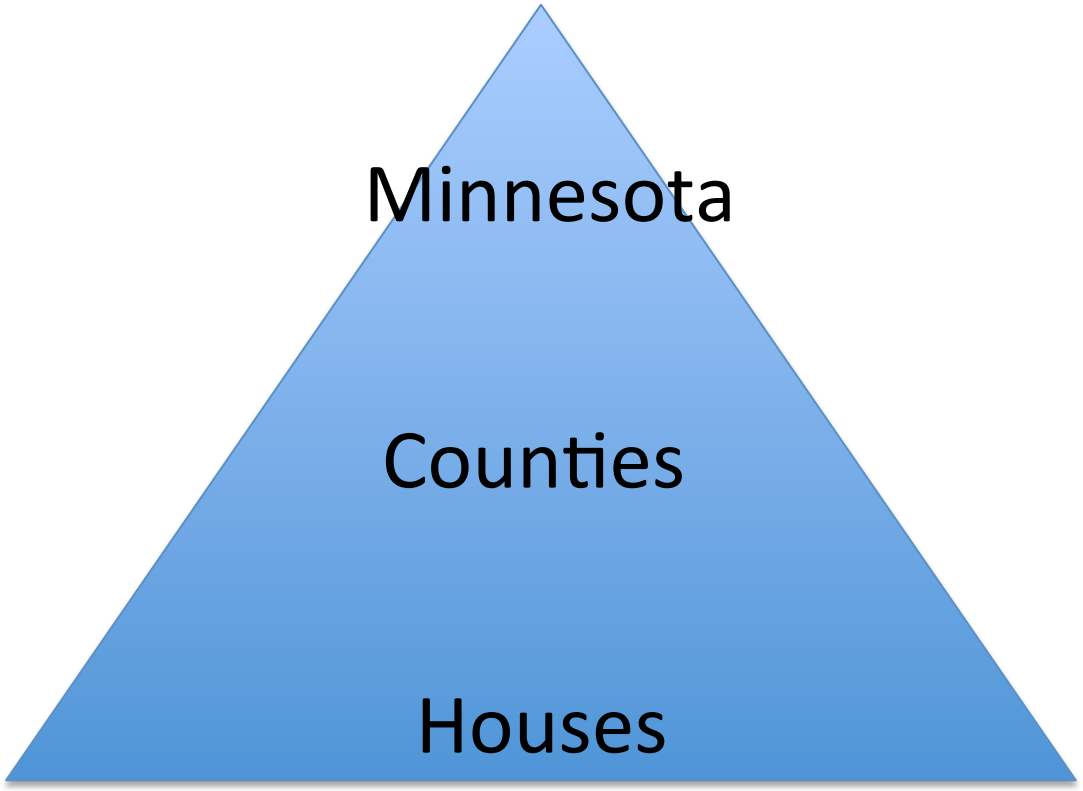
Question: do you want to live in Minnesota?

Except that this ignores differences between counties.

Pooled: all of Minnesota



Hierarchical models



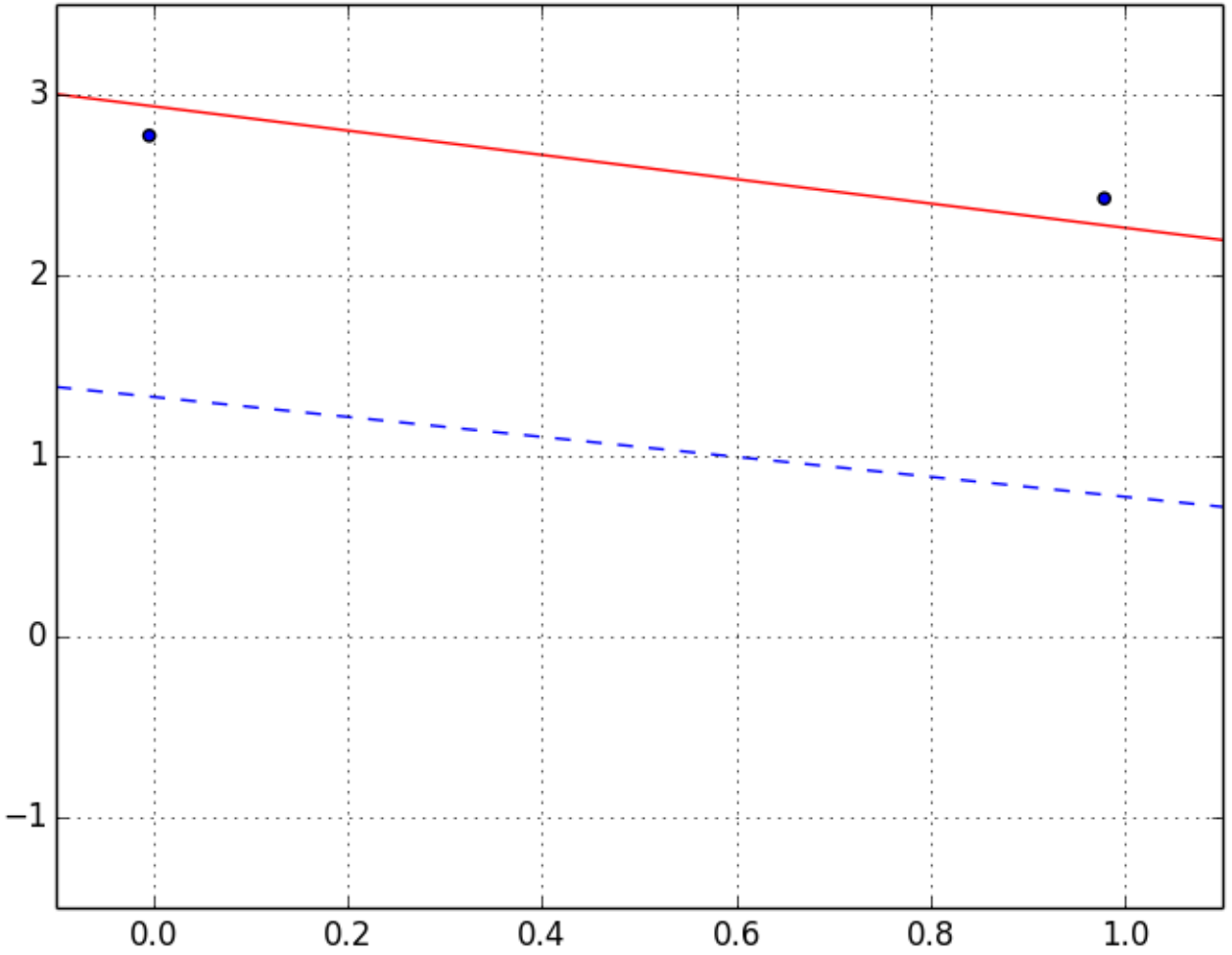
Unpooled: each county

$$radon_j = \beta_{j0} + \beta_{j1} basement?? + \epsilon$$

Question: do you want to live in Aitkin / Lac Qui Parle / St. Louis / wherever?

Different averages / badness-of-having-a-basement in each county.

Unpooled: small samples



Partial pooling: combination

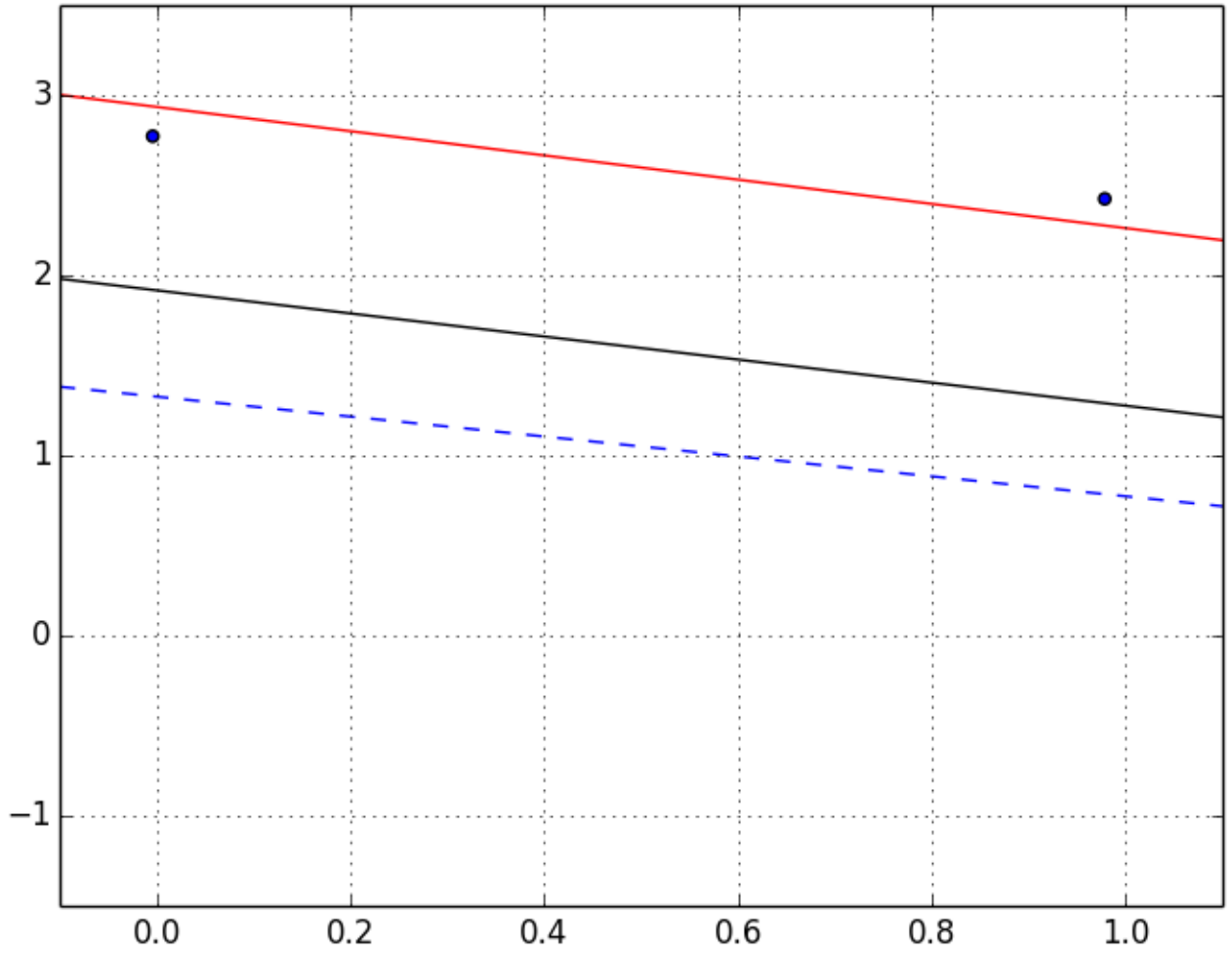
$$radon = a_{county} + \beta\ basement?? + \epsilon_{houses}$$

$$a_{county} = u + \epsilon_{counties}$$

Mixture of the two. Estimate two error terms: one for counties, and one for houses.

Not a linear problem, need to use computer-aided methods.

Partial pooling: combination



Other uses

- “Shrinkage estimation” moving group means towards an overall mean. A sample mean is a special case of linear regression.
- Education policy: students < classes/teachers < school
- Political science: voters < states < U.S.A.