

Multilevel pooled radon model

Names: (signatures only please, printed names will not be counted)

- | | |
|-----|-----|
| 1.) | 4.) |
| 2.) | 5.) |
| 3.) | 6.) |

Overview

In this exercise we will examine the Radon data using a completely pooled model, that is, we lump all of the counties together and pretend they all behave the same with respect to radon concentrations

The pooled model is:

$$y = \beta_0 + \beta_1 \cdot x + e$$

Where:

- y is the measured radon concentration
- β_0 is the intercept of the regression line
- β_1 is the coefficient of the floor (0=basement, 1=first)
- e is the noise term, assumed to be independent normal $(0, \sigma_e)$

The data elements are:

- `county` County number
- `floor_measure` Floor - 0 or 1
- `log_radon` natural log of radon concentration

While this model is nominally a simple regression, because x takes only values zero and one, it is computationally identical to a single-factor ANOVA with two levels.

The difference is that we might use this model to predict radon on the second floor by coding $x = 2$, something that would have no meaning in an ANOVA model.

The line between regression and ANOVA can be rather blurry. This model would fall under a class of models known as "regression on dummy variables" where "dummy variables" refers to predictors that only assume the

values zero and one. This terminology is common in many fields, including economics.

Like the single-factor ANOVA with two levels, there are only two fitted values, which correspond to the mean radon level for basements and first floors.

The .Rnw file is:

`Multilevel_pooled.Rnw`

Questions:

- 1) Generally you need to validate the predicted values from a model against real-world measurements. Can you see a potential problem with the radon levels produced by this model in that regard?