Lab #10 - Logistic Regression Part II

Econ 224

September 25th, 2018

Contaminated Wells in Bangladesh

Today we'll work with a dataset containing household-level information from Bangladesh: `wells.csv`. You can download the dataset from the course website at http://ditraglia.com/econ224/wells.csv.

Here is some background on the dataset from Gelman and Hill (2007):

> Many of the wells used for drinking water in Bangladesh and other South Asian countries are contaminated with natural arsenic ... a research team from the United States and Bangladesh measured all the wells [in a small region] and labeled them with their arsenic level as well as a characterization of "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 encourage 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.

Our goal is to predict which households will switch wells using the following information:

Name	Description
`dist`	Distance to closest known safe well (meters)
`arsenic`	Arsenic level of respondent's well (100s of micrograms/liter)
`switch`	Dummy variable: equals 1 if switched to a new well
`assoc`	Dummy variable: equals 1 if any member of the household is active in community organizations
`educ`	Education level of head of household (years)