Coursework 1: STAT 570

Philip Pham

October 2, 2018

1. The data we analyze are from a 1970s study that investigated insurance redlining on $n = 47$ zipcodes. Information on who was being refused homeowners is not available so instead we take as response the number of FAIR plan policies written and renewed in Chicago by zip code over the period December 1977 to May 1978. The FAIR plan was offered by the city of Chicago as a default policy to homeowners who had been rejected by the voluntary market. The data we will analyze are named `chredlin` and are in the `faraway` package. The variable `involact` are the number of new FAIR plan policies and renewals per 100 housing units.

   We will consider five covariates for modeling the response: racial composition in percent minority (`race` x_{i1}), fires per 100 housing units (`fire` x_{i2}), theft per 1000 population (`theft` x_{i3}), percent of housing units built before 1939 (`age` x_{i4}), log median family income in thousands of dollars (`lincome` x_{i5}‘), $i = 1, \ldots, 47$.

   We will examine the model with the main effects due to race, fire, theft, age and log(income).

   We let Y_i represent `involact`, and $x_i = (x_{i1}, x_{i2}, \ldots, x_{i5})$, the covariates, for individual i, $i = 1, 2, \ldots, 47$. We fit the model

$$y_i = \beta_0 + \sum_{j=1}^{5} x_{ij}\beta_j + \epsilon_i \tag{1}$$

   for $i = 1, \ldots, n$ using least squares.

   (a) Provide informative plots to illustrate what we might expect to learn from the model in Equation 1.

   Solution: See Figure 1.

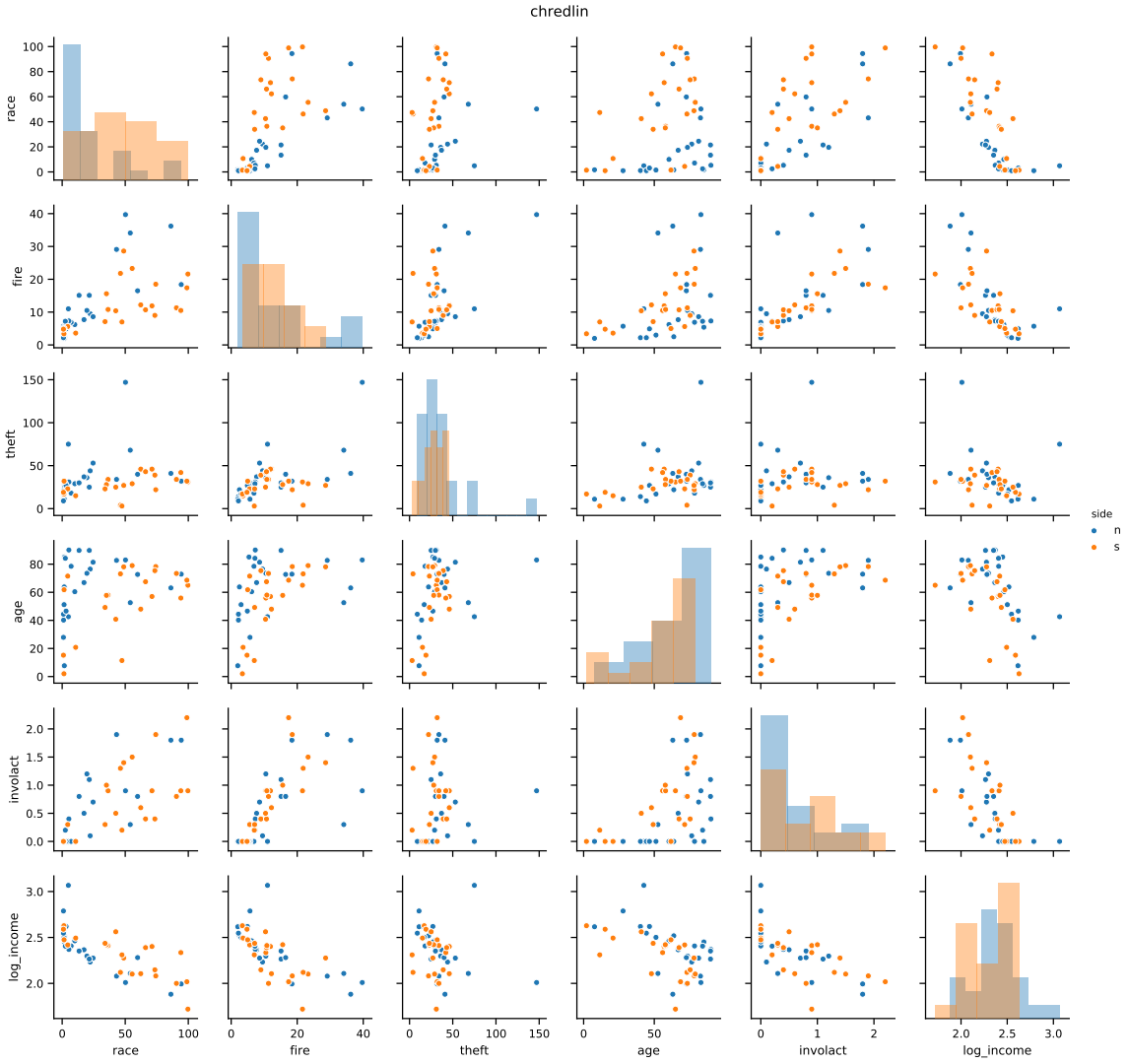

Figure 1: The empirical univariate and joint distributions for the `chredlin` dataset.