A case study

ST552 Lecture 27

Charlotte Wickham

2019-03-13

Announcements

Friday lecture: office hours instead

Can you use residual plots to look for correlated errors?

It might show up in the plots we've talked about. . .

There are some specific purpose plots too.

Follow Faraway Chapter 12

.. illustrates several of the ambiguities and difficulties encountered in statistical practice. – Faraway

Indeed, it has been said democracy is the worst form of Government except all those other forms that have been tried from time to time. – Winston Churchill

We might say the same about statistics with respect to how it helps us reason in the face of uncertainty. It is not entirely satisfying but the alternatives are worse. – Faraway

Background

> *Insurance redlining refers to the practice of refusing to issue insurance to certain types of people or within some geographic area.*

Seems reasonable to refuse someone based on their past behavior (e.g. drunk driving), but it would be wrong to refuse someone based on their race.

It is the 1970s. Chicago neighborhoods are being red-lined. Residents claim racial discrimination. Insurance company claims it's based on historical losses. (**This is a simplification**)

Available data

We don't actually have data on insurance refusals for individuals.

We do have the number of applications for FAIR plans (a city plan for people who get refused insurance) at a zip code level.

```
data(chredlin, package = "faraway")
head(chredlin)
```

```
##       race fire theft  age involact income side
## 60626 10.0  6.2    29 60.4       0.0 11.744    n
## 60640 22.2  9.5    44 76.5       0.1  9.323    n
## 60613 19.6 10.5    36 73.5       1.2  9.948    n
## 60657 17.3  7.7    37 66.9       0.5 10.656    n
## 60614 24.5  8.6    53 81.4       0.7  9.730    n
## 60610 54.0 34.1    68 52.6       0.3  8.231    n
```

Chapter 12

Insurance Redlining — A Complete Example

In this chapter, we present a relatively complete data analysis. The example is interesting because it illustrates several of the ambiguities and difficulties encountered in statistical practice.

Insurance redlining refers to the practice of refusing to issue insurance to certain types of people or within some geographic area. The name comes from the act of drawing a red line around an area on a map. Now few would quibble with an insurance company refusing to sell auto insurance to a frequent drunk driver, but other forms of discrimination would be unacceptable.

In the late 1970s, the US Commission on Civil Rights examined charges by several Chicago community organizations that insurance companies were redlining their neighborhoods. Because comprehensive information about individuals being refused homeowners insurance was not available, the number of FAIR plan policies written and renewed in Chicago by zip code for the months of December 1977 through May 1978 was recorded. The FAIR plan was offered by the city of Chicago as a default policy to homeowners who had been rejected by the voluntary market. Information on other variables that might affect insurance writing such as fire and theft rates was also collected at the zip code level. The variables are:

`race` racial composition in percentage of minority

`fire` fires per 100 housing units

`theft` thefts per 1000 population

`age` percentage of housing units built before 1939

`involact` new FAIR plan policies and renewals per 100 housing units

`income` median family income in thousands of dollars

`side` north or south side of Chicago

The data come from Andrews and Herzberg (1985) where more details of the variables and the background are provided.

12.1 Ecological Correlation

Notice that we do not know the races of those denied insurance. We only know the racial composition in the corresponding zip code. This is an important difficulty that needs to be considered before starting the analysis.

When data are collected at the group level, we may observe a correlation between two variables. The ecological fallacy is concluding that the same correlation holds

Figure 1:

Getting started

- Do some basic exploration of the data.
- Verify the observation: "Zips with higher proportions of the minority have higher rates of FAIR plan application"
- What would the insurance company argue? How could we address their argument?