536 Homework 3

Adrian Dobra adobra@uw.edu

This homework focuses on the police stops data. Start by downloading the data file "policestops.txt" from our discussion board. The variables are as follows:

- The outcome is "stops" and represents the number of police stops of members of an ethnic group in a precinct.
- The exposure is "arrests" and represents the number of arrests of people of an ethnic group in a precinct.
- The first explanatory variable is "precinct". There are 75 precincts coded from 1 to 75.
- The second explanatory variable is "ethnicity": 1 means black, 2 means hispanic and 3 means white.

You need to study whether there is any evidence that the police tends to stop people of a certain ethnicity more often than people of some other ethnicity. Follow these steps:

(A) Fit the Poisson models

$$\begin{aligned}
\text{stops} &= \text{Poisson}(\exp(\beta_0 + \beta_1 \cdot \text{precinct} + \beta_2 \cdot \text{ethnicity})),\\
\text{stops} &= \text{Poisson}(\exp(\beta_0 + \beta_1 \cdot \text{precinct})),\\
\text{stops} &= \text{Poisson}(\exp(\beta_0 + \beta_2 \cdot \text{ethnicity})),\\
\text{stops} &= \text{Poisson}(\exp(\beta_0)).
\end{aligned}$$

An example of fitting one of these models in R is:

```
glm(stops ~ factor(ethnicity)+factor(precinct),family=poisson(link = "log"))
```

Which model would you prefer based on AIC? Which model would you prefer based on BIC? Given your choice of model, what is the expected number of stops in precinct 3 for blacks? What is the expected number of stops in precinct 3 for whites? Answer the same questions for precinct 1.

(B) Now use the number of arrests as baseline. That is, fit the following models:

$$\begin{aligned}
\text{stops} &= \text{Poisson}(\text{arrests} \cdot \exp(\beta_0 + \beta_1 \cdot \text{precinct} + \beta_2 \cdot \text{ethnicity})),\\
\text{stops} &= \text{Poisson}(\text{arrests} \cdot \exp(\beta_0 + \beta_1 \cdot \text{precinct})),\\
\text{stops} &= \text{Poisson}(\text{arrests} \cdot \exp(\beta_0 + \beta_2 \cdot \text{ethnicity})),\\
\text{stops} &= \text{Poisson}(\text{arrests} \cdot \exp(\beta_0)).
\end{aligned}$$

An example of fitting one of these models in R is:

```
glm(stops ~ factor(ethnicity)+factor(precinct),family=poisson(link = "log"),offset=log(arrests))
```

Which model would you prefer based on AIC? Which model would you prefer based on BIC? Would you prefer a model with a baseline or a model without a baseline? Justify your answers. Find out the expected number of stops in precincts 1 and 3 for blacks and whites based on your final choice of model.

(C) Intepret the coefficients of your final model. Is there any evidence of racial discrimination by the police?