536HW3

Coco_Luo

2022-10-24

Question A

We fit 4 models

$$M_0 : stops = Poisson(exp(\beta_0 + \beta_1 \cdot myprecinct + \beta_2 \cdot myethnicity))$$

The first model has AIC: 56347.60, BIC: 54904.03.

$$M_1 : stops = Poisson(exp(\beta_0 + \beta_1 \cdot myprecinct))$$

The second model has AIC: 90905.00, BIC: 89454.59.

$$M_2 : stops = Poisson(exp(\beta_0 + \beta_2 \cdot myethnicity))$$

The third model has AIC: 90483.73, BIC: 88787.37.

$$M_3 : stops = Poisson(exp(\beta_0))$$

The fourth model has AIC: 125041.13, BIC: 123337.93.

	AICs	BICs
model1	56347.60	54904.03
model2	90905.00	89454.59
model3	90483.73	88787.37
model4	125041.13	123337.93

From the results above we can see that M_0 has the smallest AIC and BIC so I would choose M_0 than other models.

The baseline for "ethnicity" is black and precinct 1 is the baseline for "precinct". The expected number of police stops in precinct 3 for black is:

$$E[stops|myprecinct = 3, myethnicity = 1] = e^{\beta_0 + \beta_{pre=3}} = 851.67$$

The expected number of police stops in precinct 3 for white is:

$$E[stops|myprecinct = 3, myethnicity = 3] = e^{\beta_0 + \beta_{pre=3} + \beta_{eth=3}} = 207.04$$

The expected number of police stops in precinct 1 for black is:

$$E[stops|myprecinct = 1, myethnicity = 1] = e^{\beta_0} = 204.55$$

The expected number of police stops in precinct 1 for white is:

$$E[stops|myprecinct = 1, myethnicity = 3] = e^{\beta_0 + \beta_{eth=3}} = 49.73$$

	prediction
precinct = 3 and ethnicity = 1	851.67
precinct = 3 and ethnicity = 3	207.04
precinct = 1 and ethnicity = 1	204.55
precinct = 1 and ethnicity = 3	49.73

Question B

We fit 4 models

$$B_0 : stops = Poisson(arrests \cdot exp(\beta_0 + \beta_1 \cdot myprecinct + \beta_2 \cdot myethnicity))$$

The first model has AIC: 5287.752, BIC: 3844.179

$$B_1 : stops = Poisson(arrests \cdot exp(\beta_0 + \beta_1 \cdot myprecinct))$$

The second model has AIC: 7752.906, BIC: 6302.501

$$B_2 : stops = Poisson(arrests \cdot exp(\beta_0 + \beta_2 \cdot myethnicity))$$

The third model has AIC: 47149.965, BIC: 45453.601.

$$B_3 : stops = Poisson(arrests \cdot exp(\beta_0))$$

The fourth model has AIC: 47828.875, BIC: 46125.679.

Again, the first model has the smallest AIC and BIC, and it has the smallest scores among all 8 models we have fitted so far. I prefer the model with a baseline since it adjusted the counts of `stops` to the ratio of `stops` and `arrests` so the predicted values are comparable.

	AICs	BICs
modelB1	5287.752	3844.179
modelB2	7752.906	6302.501
modelB3	47149.965	45453.601
modelB4	47828.875	46125.679

We use the number of arrests as baseline

The baseline for "ethnicity" is black and precinct 1 is the baseline for "precinct". The expected number of police stops in precinct 3 for black is:

$$E[stops|myprecinct = 3, myethnicity = 1] = arrests_{pre=3,eth=1} \cdot e^{\beta_0+\beta_{pre=3}} = 964.71$$

The expected number of police stops in precinct 3 for white is:

$$E[stops|myprecinct = 3, myethnicity = 3] = arrests_{pre=3,eth=3} \cdot e^{\beta_0+\beta_{pre=3}+\beta_{eth=3}} = 359.01$$

The expected number of police stops in precinct 1 for black is:

$$E[stops|myprecinct = 1, myethnicity = 1] = arrests_{pre=1,eth=1} \cdot e^{\beta_0} = 246.83$$

The expected number of police stops in precinct 1 for white is:

$$E[stops|myprecinct = 1, myethnicity = 3] = arrests_{pre=1,eth=3} \cdot e^{\beta_0+\beta_{eth=3}} = 63.11$$

	predictions
precinct = 3 and ethnicity = 1	964.71
precinct = 3 and ethnicity = 3	359.01
precinct = 1 and ethnicity = 1	246.83
precinct = 1 and ethnicity = 3	63.11

Question C

The final model I choose to fit my data is modelB1: stops = Poisson(arrests $\cdot$ exp(β_0 + $\beta_1\cdot$ precinct + $\beta_2\cdot$ ethnicity)), where black and myprecinct 1 are the baseline categories for `myeth` and `myprecinct`. This means that holding all other variables constant, for each unit of change in $\frac{stops_{i,k}}{arrests_{i,k}}$, the expected count changes by a factor of exp(β_i^k).

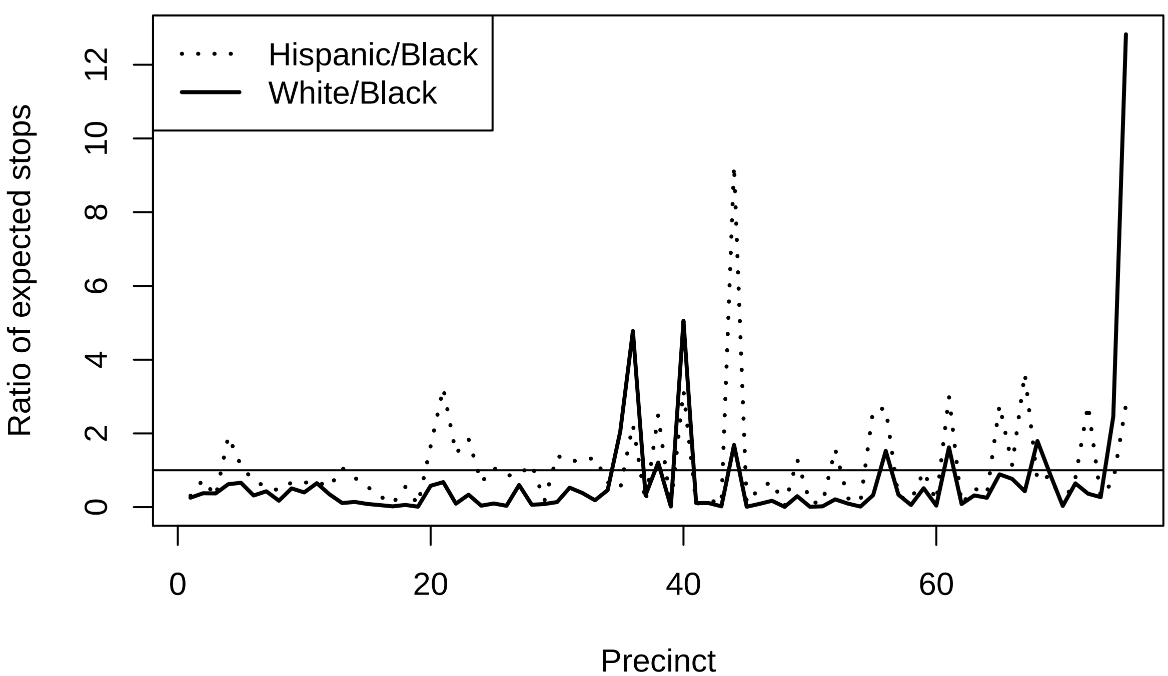
And under the chosen model, we have the following relationships

$$\frac{E[stops|myprecinct, myethnicity = 2]}{E[stops|myprecinct, myethnicity = 1]} = e^{\beta_{eth=2}} = 0.64$$

$$\frac{E[stops|myprecinct, myethnicity = 3]}{E[stops|myprecinct, myethnicity = 1]} = e^{\beta_{eth=3}} = 0.24$$

Which tells us that holding all other variables constant, in each of the 75 precincts, the expected number of police stops for Hispanices is about 64% the expected number of polices stops for blacks, and the expected number of police stops for whites is about 24% the expected number of polices stops for blacks. Thus, there is racial discrimination against the black and Hispanic populations. Since the number of arrests may vary by precinct and ethnicity, we would expect the ratios of the expected number of stops to be different among precincts.

$$\frac{E[stops|myprecinct, myethnicity = 2]}{E[stops|myprecinct, myethnicity = 1]} = \frac{arrests_{pre,eth=2}}{arrests_{pre,eth=1}} e^{\beta_{eth=2}} = 0.64$$
$$\frac{E[stops|myprecinct, myethnicity = 3]}{E[stops|myprecinct, myethnicity = 1]} = \frac{arrests_{pre,eth=3}}{arrests_{pre,eth=1}} e^{\beta_{eth=3}} = 0.64$$



The black solid line represents the ratio of the expected number of stops for hispanics and the expected number of stops for blacks, while the dotted line represents the ratio of the expected number of stops for whites and the expected number of stops for blacks. Ratios of expected number of stops below the horizontal line y = 1 indicates fewer stops for hispanics or whites with respect to blacks, and above y=1 means more stops. In general, we can see that blacks tend to be stopped more frequently than the other two echnicities since both curves stay below y=1 for most of the precinct. The dotted line have more areas above the line than the solid line, showing that more hispanics are stoped compared to white. As a result, I confirmed that there is evidence of racial discrimination by the police.

Appendix - Code

```
model1 = glm(mystops ~ factor(myeth) + factor(myprecinct), data = police, family=poisson(link = "log"))
model2 = glm(mystops ~ factor(myprecinct), data = police, family=poisson(link = "log"))
model3 = glm(mystops ~ factor(myeth),  data = police, family=poisson(link = "log"))
model4 = glm(mystops ~ 1,  data = police, family=poisson(link = "log"))
```

```
AICs = rep(0, 4)
AICs = AIC(model1, model2, model3, model4)$AIC

b1 = model1$deviance+length(coef(model1))*log(dim(police)[[1]])
b2 = model2$deviance+length(coef(model2))*log(dim(police)[[1]])
b3 = model3$deviance+length(coef(model3))*log(dim(police)[[1]])
b4 = model4$deviance+length(coef(model4))*log(dim(police)[[1]])

BICs = rep(0, 4)
BICs = c(b1,b2,b3,b4)

tableA = data.frame(AICs, BICs, row.names = c("model1", "model2", "model3", "model4"))
knitr::kable(tableA)
```

```
test = data.frame(myprecinct = c(3, 3, 1, 1), myeth = c(1, 3, 1, 3))
prediction = round(predict(model1, test, type = "response"), 2)
tableB = data.frame(prediction, row.names = c("model1", "model2", "model3", "model4"))
knitr::kable(tableB)
```

```
modelB1 = glm(mystops ~ factor(myeth) + factor(myprecinct),
              data = police, family=poisson(link = "log"), offset=log(myarrests))
modelB2 = glm(mystops ~ factor(myprecinct),
              data = police, family=poisson(link = "log"), offset=log(myarrests))
modelB3 = glm(mystops ~ factor(myeth),
              data = police, family=poisson(link = "log"), offset=log(myarrests))
modelB4 = glm(mystops ~ 1,
              data = police, family=poisson(link = "log"), offset=log(myarrests))

AICs = rep(0, 4)
AICs = AIC(modelB1, modelB2, modelB3, modelB4)$AIC

b_1 = modelB1$deviance+length(coef(modelB1))*log(dim(police)[[1]])
b_2 = modelB2$deviance+length(coef(modelB2))*log(dim(police)[[1]])
b_3 = modelB3$deviance+length(coef(modelB3))*log(dim(police)[[1]])
b_4 = modelB4$deviance+length(coef(modelB4))*log(dim(police)[[1]])

BIC_s = rep(0, 4)
BIC_s = c(b_1,b_2,b_3,b_4)
```

```
tableC = data.frame(AICs, BIC_s, row.names = c("modelB1", "modelB2", "modelB3", "modelB4"))
knitr::kable(tableC)
```

```
test = data.frame(myprecinct = c(3, 3, 1, 1),
                  myeth = c(1, 3, 1, 3),
                  myarrests = c(2188, 1238, 980, 381))
predictions = round(predict(modelB1, test, type = "response"), 2)
tableB0 = data.frame(predictions, row.names = c("precinct = 3 and ethnicity = 1", "precinct = 3 and ethn
knitr::kable(tableB0)
```

```
arrests = police$myarrests
precinct = factor(police$myprecinct)
ethnicity = factor(police$myeth)

ratio_h = rep(0,75)
ratio_w = rep(0,75)

for(i in 1:75)
{
 ratio_h[i] = arrests[(precinct == i)&(ethnicity == 2)] * exp(coef(modelB1)["factor(myeth)2"])/
              arrests[(precinct ==i)&(ethnicity==1)]
 ratio_w[i] = arrests[(precinct == i)&(ethnicity == 3)] * exp(coef(modelB1)["factor(myeth)3"])/
              arrests[(precinct==i)&(ethnicity==1)]
}

plot(1:75, ratio_w, type="l", lwd=2,
     xlab = "Precinct", ylab="Ratio of expected stops")
lines(1:75, ratio_h, lwd=2, lty="dotted")
abline(h = 1)
legend(legend = c("Hispanic/Black", "White/Black"), "topleft",
       lty = c("dotted", "solid"), lwd = c(2,2))
```