Data Incubator

Project Proposal

Elham Keshavarzi

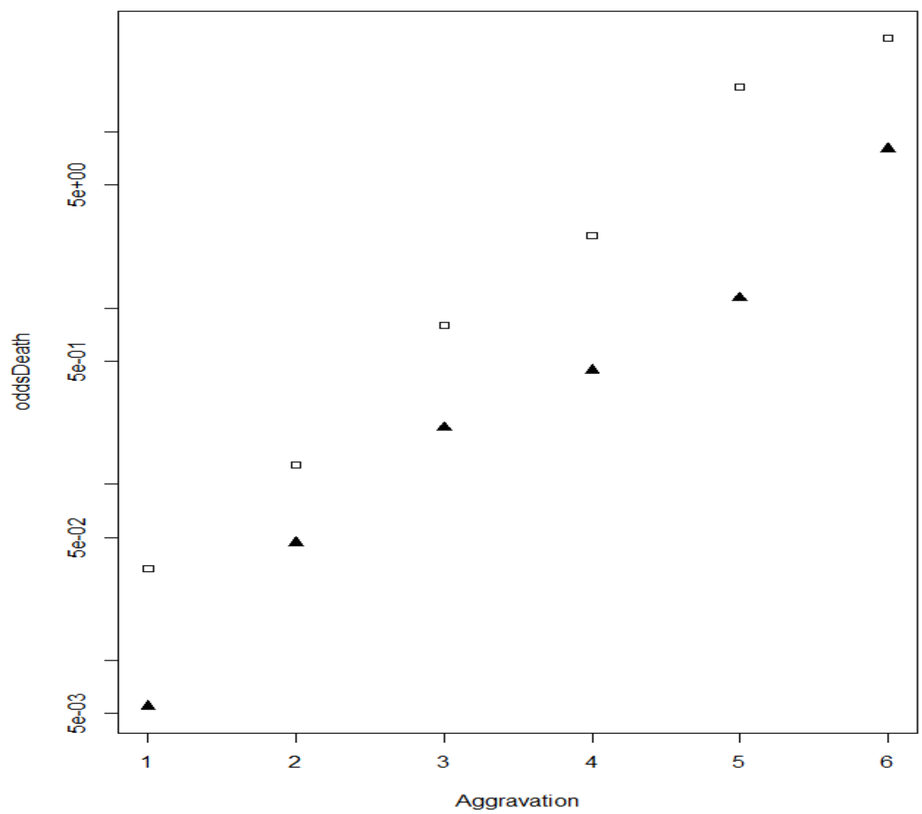
Problem

Racism is still a controversial issue! In this project I'm studying the racism in criminal accusations. The highest level of punishment for a crime is death penalty. In this project I investigate the data related to african-american criminals and white criminals who got death penalty for commiting a murder. My question is if there is a meaningful difference between getting the death penalty between black and whites.

To complete this project I applied data available in R library, Sleuth3 package. To answer the question I applied a logistic regression method to fit a model to data and test the significant of race and interpret the results.

Data visualization

The initial plot of data suggests a relationship between empirical logits for the death penalty and race of victim. The plot suggests there is a meaningful difference for the black victim murderers and the white victim murderers.



The logistic regression model that I fit to the odds of death for white victims accounting the aggravation level is:

$$logit(\hat{\pi}) = -6.2319 \ + 1.4067(\text{AggravationLevel}) + 0.9788(White\ Victim) + 0.2744(\text{AggravationLevel})\ (White\ Victim)$$

The p-values for aggravation is 1.17e-08 which is really small and shows that this variable is significant, however the p-value for the race 0.429 and p-value for the interaction between race and aggravation 0.466 is large, which shows that the indicator is not significant. The residual deviance for this model is 3.3438 and AIC is 33.21. Follow is the table of all the coefficient from the fitting the full model.

```
Coefficients:
                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)             -6.2319     0.9015  -6.913 4.75e-12 ***
Aggravation              1.4067     0.2466   5.705 1.17e-08 ***
VictimWhite              0.9788     1.2377   0.791    0.429    
Aggravation:VictimWhite  0.2744     0.3762   0.729    0.466    
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.3438  on  8  degrees of freedom
AIC: 33.21
```

Check the importance of multiplied term:

The logistic regression model that I fit to the odds of death for white victims accounting the aggravation level is:

$$logit(\hat{\pi}) = -6.6760 + 1.5397(\text{AggravationLevel}) + 1.8106(White\ Victim)$$

The p-values for aggravation is 2e-16 which is really small and shows that this variable is significant, also the p-value for the race (White victims) is small 0.000732 shows the significant of white race. The residual deviance for this model is 3.8816 and AIC is 31.747. This model without the interaction between race and aggravation is a better fit because of lower deviance and lower AIC. Follow is the table of all the coefficient from the model without race interaction.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -6.6760     0.7574  -8.814  < 2e-16 ***
Aggravation   1.5397     0.1867   8.246  < 2e-16 ***
Victimwhite   1.8106     0.5361   3.377 0.000732 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.8816  on  9  degrees of freedom
AIC: 31.747
```

Compare two models:

The analysis of deviance shows the same result that the model without the indicator is a better fit because it provides 0.53778 less deviance. Here is the comparison table:

```
Model 1: cbind(Death, NoDeath) ~ Aggravation + Victim + Aggravation:Victim
Model 2: cbind(Death, NoDeath) ~ Aggravation + Victim
  Resid. Df Resid. Dev Df Deviance
1         8     3.3438            
2         9     3.8816 -1 -0.53778
```

Test the significant of race

The estimated coefficient for the race is 1.8106 and the z-statistic is 3.377. The null hypothesis is that the coefficient is 0 and the alternative is that the coefficient is not equal to zero. The wald's test has normal distribution under the null hypothesis. The two sided p-value is 0.000732 which is small and we reject the null hypothesis and the conclusion is that based on Wald's test coefficient is not equal to zero for the victim race and being white or black matters to get the death penalty.

$$H0: \beta_2 = 0$$

$$H0: \beta_2 \neq 0$$

z-statistic=3.377

Two sided p-value = 0.0007328106

Standard Error = 0.5361

%95 confidance interval

exp[1.8106 -1.96*0.5361   1.8106 +1.96*0.5361]= [2.137943   17.48522]

With 95% confidence the odds of getting death penalty for the white victim murderers to the odds of getting death penalty for the black victim murderers after accounting for aggravation level of the crime is in the range [2.137943   17.48522].

So the racism exists at least in our courts!

R codes

```
> install.packages("Sleuth3")
Installing package into ‘C:/Users/Alice/Documents/R/win-library/3.3’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/Sleuth3_1.0-2.zi
p'
Content type 'application/zip' length 4417981 bytes (4.2 MB)
downloaded 4.2 MB

package ‘Sleuth3’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Alice\AppData\Local\Temp\RtmpUxZ3B5\downloaded_packages
> library(Sleuth3)
> attach(case1902)
> names(case1902)
[1] "Aggravation" "Victim"      "Death"       "NoDeath"
> case1902
   Aggravation Victim Death NoDeath
1            1  White     2      60
2            1  Black     1     181
```

```
3             2  White     2      15
4             2  Black     1      21
5             3  White     6       7
6             3  Black     2       9
7             4  White     9       3
8             4  Black     2       4
9             5  White     9       0
10            5  Black     4       3
11            6  White    17       0
12            6  Black     4       0
> deathProp <- Death/(Death + NoDeath)
> oddsDeath <- Death/(NoDeath +0.5)
> shapePlot <- ifelse(Victim=="White",22,24)
> colorPlot <- ifelse(Victim=="White","white","black")
> plot(oddsDeath ~ Aggravation, log="y", pch=shapePlot, bg=colorPlot)
> logModelWithRace <- glm(cbind(Death,NoDeath) ~ Aggravation + Victim +
+                 Aggravation:Victim, family=binomial)
> summary(logModelWithRace)

Call:
glm(formula = cbind(Death, NoDeath) ~ Aggravation + Victim +
    Aggravation:Victim, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6614  -0.2246   0.1658   0.5430   0.9132

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -6.2319     0.9015  -6.913 4.75e-12 ***
Aggravation               1.4067     0.2466   5.705 1.17e-08 ***
VictimWhite               0.9788     1.2377   0.791    0.429
Aggravation:VictimWhite   0.2744     0.3762   0.729    0.466
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.3438  on  8  degrees of freedom
AIC: 33.21

Number of Fisher Scoring iterations: 4

> logModelWithoutRace <- update(logModelWithRace, ~ . - Aggravation:Victim)
> summary(logModelWithoutRace)

Call:
glm(formula = cbind(Death, NoDeath) ~ Aggravation + Victim, family = binomial
)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93570  -0.22548   0.05142   0.65620   1.01444

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)  -6.6760     0.7574  -8.814  < 2e-16 ***
Aggravation   1.5397     0.1867   8.246  < 2e-16 ***
VictimWhite   1.8106     0.5361   3.377 0.000732 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.8816  on  9  degrees of freedom
AIC: 31.747

Number of Fisher Scoring iterations: 4

> anova(logModelWithRace, logModelWithoutRace) # no evidence of interaction.
Analysis of Deviance Table

Model 1: cbind(Death, NoDeath) ~ Aggravation + Victim + Aggravation:Victim
Model 2: cbind(Death, NoDeath) ~ Aggravation + Victim
  Resid. Df Resid. Dev Df Deviance
1         8     3.3438
2         9     3.8816 -1 -0.53778
> fitGoodness <- 1-pchisq(logModelWithoutRace$deviance,df=logModelWithoutRace
$df.residual)
> fitGoodness
[1] 0.9190319
```