Hw6

Trevor Freeland

May 1, 2018

```r
library(tidyverse)
library(pander)
```

```r
prison <- read.csv("http://math.carleton.edu/Chihara/Stats345/Recidivism.csv")
```

54

(A)

$f(y; \lambda) = exp[rlog(\lambda) + (r-1)log(y) - log(\Gamma(r) - y\lambda]$

$f(y; \lambda) = exp[-y\lambda + rlog(\lambda) + (r-1)log(y) - log(\Gamma(r)]$

If we let $\Theta = \lambda$ then we can see this is from the exponential family.

(B)

E(a(Y)) $= \frac{-c'(\lambda)}{d'(\lambda)}$

E(a(Y)) $= \frac{\delta}{\delta\lambda} * \frac{-(rlog(\lambda))}{(-\lambda)}$

E(a(Y)) $= \frac{r}{\lambda}$

Var(a(Y)) $= \frac{b''(\lambda)c'(\lambda) - c''(\lambda)d'(\lambda)}{b'(\lambda)^3}$

Var(a(Y)) $= \frac{0 + \frac{r}{\lambda^2} * -1}{-1}$

Var(a(Y)) $= \frac{r}{\lambda^2}$

55

(A)

$f(y; \alpha) = exp[log(\alpha) - (\alpha+1)log(y)]$ If we let $\Theta = \alpha$ then we can see this is from the exponential family.

(B)

E(log(Y)) $= \frac{-c'(\lambda)}{d'(\lambda)}$

E(log(Y)) $= \frac{-1/\alpha}{-1}$

E(log(Y)) $= \frac{1}{\alpha}$

56

(A)

In our table below we look at some of the summary statistics of our data. When looking at this table we can see that under timeserved the maximum time served is 219 months and the mean and median are both under 20, this means we could have a potential outlier or two that we might want to be weary of. Then graphically I looked at a few histograms of the data grouped by whether or not the released prisoner was back in prison. In our age plot we can see the histogram shifts slightly left which indicates that younger aged prisoners may be more likely to return to prisoner which makes some intuitive sense. Our other histogram looks at timeserved and there appears to possibly be a minor shift to the right, indicating that the longer someone serves the more likely they may be to return to prison on average.

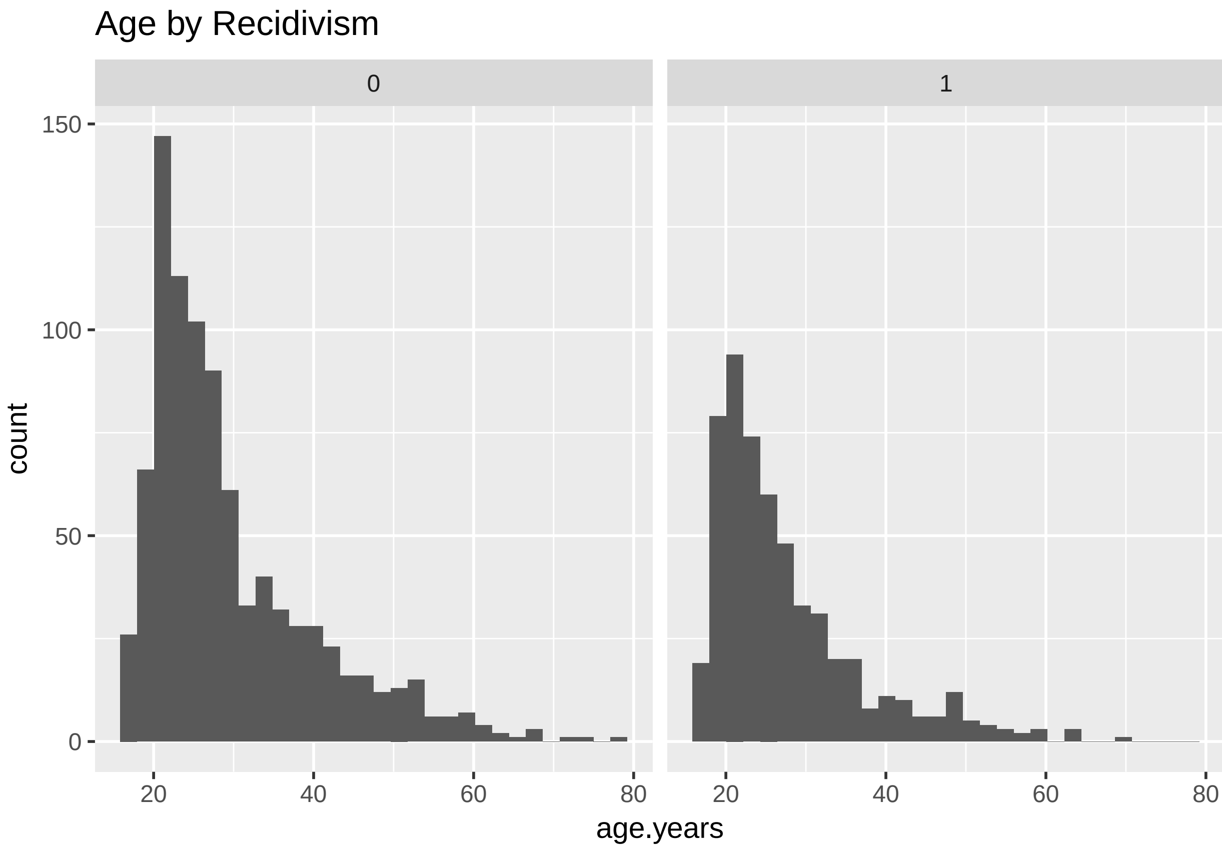
```
prison.small <- prison %>% select(c("backinprison", "alcohol", "drugs", "married", "felon", "priors", "
pander(summary(prison.small))
```

Table 1: Table continues below

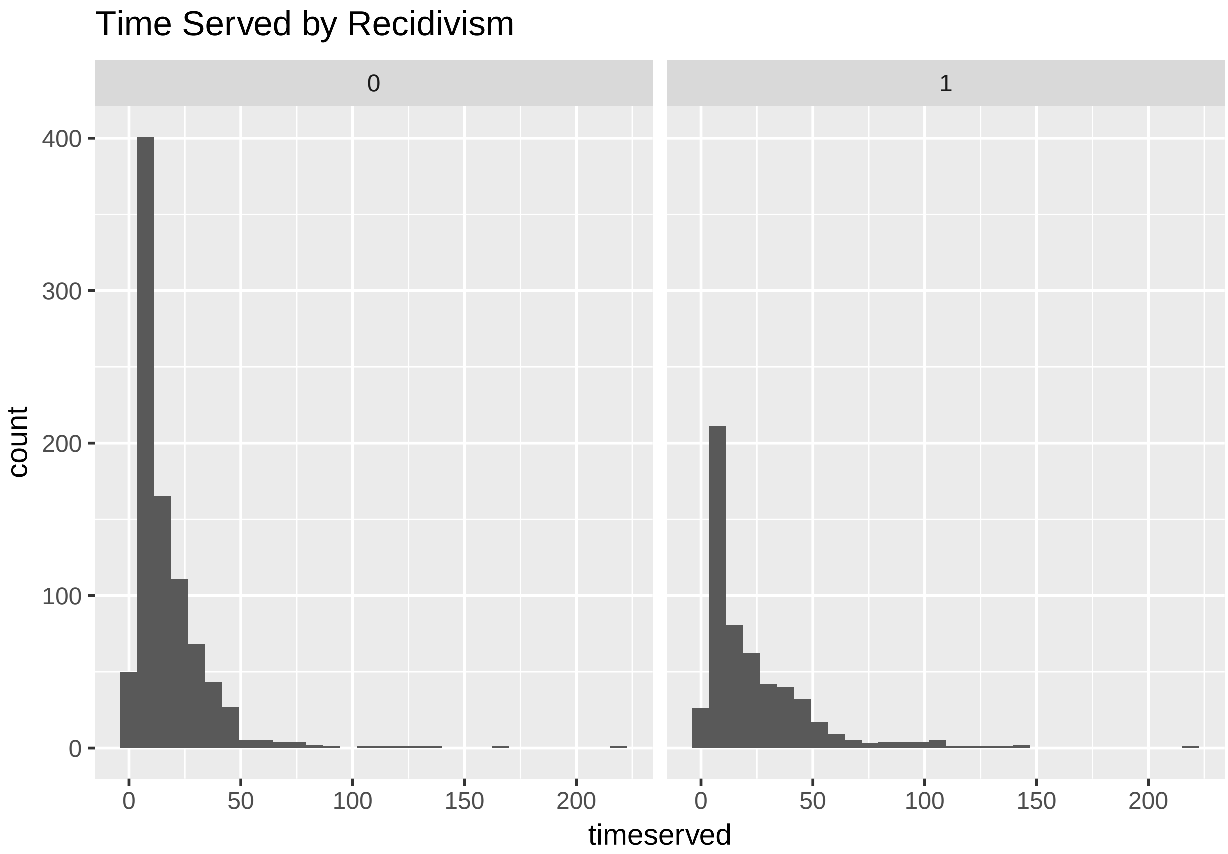
backinprison	alcohol	drugs	married
Min. :0.000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :0.000	Median :0.0000	Median :0.0000	Median :0.0000
Mean :0.382	Mean :0.2097	Mean :0.2415	Mean :0.2554
3rd Qu.:1.000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:1.0000
Max. :1.000	Max. :1.0000	Max. :1.0000	Max. :1.0000

felon	priors	educ	age.years	timeserved
Min. :0.0000	Min. : 0.000	Min. : 1.000	Min. :16.50	Min. : 0.00
1st Qu.:0.0000	1st Qu.: 0.000	1st Qu.: 8.000	1st Qu.:21.50	1st Qu.: 6.00
Median :0.0000	Median : 0.000	Median :10.000	Median :25.58	Median : 12.00
Mean :0.3142	Mean : 1.432	Mean : 9.702	Mean :28.79	Mean : 19.18
3rd Qu.:1.0000	3rd Qu.: 2.000	3rd Qu.:11.000	3rd Qu.:32.92	3rd Qu.: 25.00
Max. :1.0000	Max. :28.000	Max. :19.000	Max. :77.75	Max. :219.00

```
ggplot(prison.small, aes(x = age.years)) + geom_histogram() + facet_grid(~backinprison) + ggtitle("Age 
```




```
ggplot(prison.small, aes(x = timeserved)) + geom_histogram() + facet_grid(~backinprison) + ggtitle("Tim
```



(B)

$log(\frac{\hat{p}}{1-\hat{p}}) = .27 + .44(AlcoholYes) - .39(FelonYes) + .25(DrugsYes) + .13(Priors) - .05(Age) + .02(TimeServed)$

```
prison.lm <- glm(backinprison~alcohol+drugs+married+felon+priors+educ+age.years+timeserved, family = bi
summary(prison.lm)
prison.lm2 <- glm(backinprison~alcohol+drugs+married+felon+priors+age.years+timeserved, family = binomi
summary(prison.lm2)
anova(prison.lm,prison.lm2, test="Chisq")

#An anova taking both drugs and married out gives me a p-value of .04 so it implies one of them is not 

prison.lm3 <- glm(backinprison~alcohol+felon+drugs+priors+age.years+timeserved, family = binomial, data
summary(prison.lm3)
anova(prison.lm3,prison.lm2, test="Chisq")
```

57

(A)

The odds ratio for a 20-yr old returning to prison compared to a 30 year old returning to prison is 1.65. This indicates that the odds of a 20-yr old returning to prison is 65% larger than the odds a 30-yr old returns to prison holding everything else constant.

```
odds.ratio <- exp(-.05*20)/exp(-.05*30)
```

(B)

In our plot we show the different effects of Age between prisoners who had no priors, were not felons, spent 24 months in prison, were not convicted on drug charges and either did or did not have alcohol = 1. THe Red line is the distribution for those who had alcohol = 1, and the blue line is for the group that had alcohol = 0.

```
#Intercept + 24 months of time served + Age + Alcohol (Red)
fun1 <- function(X){
  Y <- .27  + -.05 * X + .44 + 24*.02
  p <- exp(Y)/(1+exp(Y))
  return(p)
}

#Intercept + 24 months of time served + Age (Blue)
fun2 <- function(X){
  Y <- .27  + -.05 * X  + 24*.02
  p <- exp(Y)/(1+exp(Y))
  return(p)
}
plot(0:100, fun1(0:100), type="l", col="red", ylab = "Probability of Returning to Prison" , xlab="Age (y
lines(0:100, fun2(0:100), col="blue")
```