Project3

Rongkui Han

2/6/2020

Contents

```r
library(AER)
```

```
## Warning: package 'AER' was built under R version 3.6.2

## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.6.2

## Loading required package: survival
```

```r
library(sqldf)
```

```
## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite
```

```r
library(kableExtra)
library(knitr)
library(latexpdf)
```

Make suggestions to policymakers to take certain measures by discovering variables that caused the reduction or increase of traffic fatalities.

```
data("Fatalities")
data = Fatalities
head(data)
```

```
##   state year spirits unemp   income   emppop  beertax baptist  mormon drinkage
## 1    al 1982    1.37  14.4 10544.15 50.69204 1.539379 30.3557 0.32829    19.00
## 2    al 1983    1.36  13.7 10732.80 52.14703 1.788991 30.3336 0.34341    19.00
## 3    al 1984    1.32  11.1 11108.79 54.16809 1.714286 30.3115 0.35924    19.00
## 4    al 1985    1.28   8.9 11332.63 55.27114 1.652542 30.2895 0.37579    19.67
## 5    al 1986    1.23   9.8 11661.51 56.51450 1.609907 30.2674 0.39311    21.00
## 6    al 1987    1.18   7.8 11944.00 57.50988 1.560000 30.2453 0.41123    21.00
##        dry youngdrivers    miles breath jail service fatal nfatal sfatal
## 1 25.0063     0.211572 7233.887     no   no      no   839    146     99
## 2 22.9942     0.210768 7836.348     no   no      no   930    154     98
## 3 24.0426     0.211484 8262.990     no   no      no   932    165     94
## 4 23.6339     0.211140 8726.917     no   no      no   882    146     98
## 5 23.4647     0.213400 8952.854     no   no      no  1081    172    119
## 6 23.7924     0.215527 9166.302     no   no      no  1110    181    114
##   fatal1517 nfatal1517 fatal1820 nfatal1820 fatal2124 nfatal2124   afatal
## 1        53          9        99         34       120         32 309.438
## 2        71          8       108         26       124         35 341.834
## 3        49          7       103         25       118         34 304.872
## 4        66          9       100         23       114         45 276.742
## 5        82         10       120         23       119         29 360.716
## 6        94         11       127         31       138         30 368.421
##       pop   pop1517   pop1820   pop2124 milestot unempus emppopus         gsp
## 1 3942002 208999.6 221553.4 290000.1    28516     9.7     57.8 -0.02212476
## 2 3960008 202000.1 219125.5 290000.2    31032     9.6     57.9  0.04655825
## 3 3988992 197000.0 216724.1 288000.2    32961     7.5     59.5  0.06279784
## 4 4021008 194999.7 214349.0 284000.3    35091     7.2     60.1  0.02748997
## 5 4049994 203999.9 212000.0 263000.3    36259     7.0     60.7  0.03214295
## 6 4082999 204999.8 208998.5 258999.8    37426     6.2     61.5  0.04897637
```

Question: Which factor variables change across time within each state?

```
factorchange = sqldf("SELECT state, year, breath, jail, service FROM data")

factorchange_table = data.frame(State = c("CO","CT","IL","IN","IA","KS","KY","LA","MS","NV","NH","OH","(
                                Breath = c("1982-1983","NA","1986 - 1987","1983-1984","1982-1983","1985-
                                Jail = c("NA","1984-1985","NA","NA","NA","NA","NA","NA","NA","1982-1983"
                                Service = c("NA","1984-1985","NA","NA","NA","NA","NA","NA","NA","1982-1
                                )

kable(factorchange_table)
```

State	Breath	Jail	Service
CO	1982-1983	NA	NA
CT	NA	1984-1985	1984-1985
IL	1986 - 1987	NA	NA
IN	1983-1984	NA	NA
IA	1982-1983	NA	NA
KS	1985-1986	NA	NA
KY	1983-1984	NA	NA
LA	1982-1983	NA	NA
MS	1982-1983	NA	NA
NV	NA	1982-1983	1982-1983
NH	1982-1983	NA	NA
OH	NA	1982-1983 & 1986 - 1987	NA
OR	NA	1983-1984	1983-1984
UT	NA	1982-1983	1982-1983

```r
library(ggplot2)

data = data[complete.cases(data),]
dim(data)
```

```
## [1] 335   34
```

```r
ggplot(data, aes(x = jail, y = fatal)) +
  geom_boxplot()
```
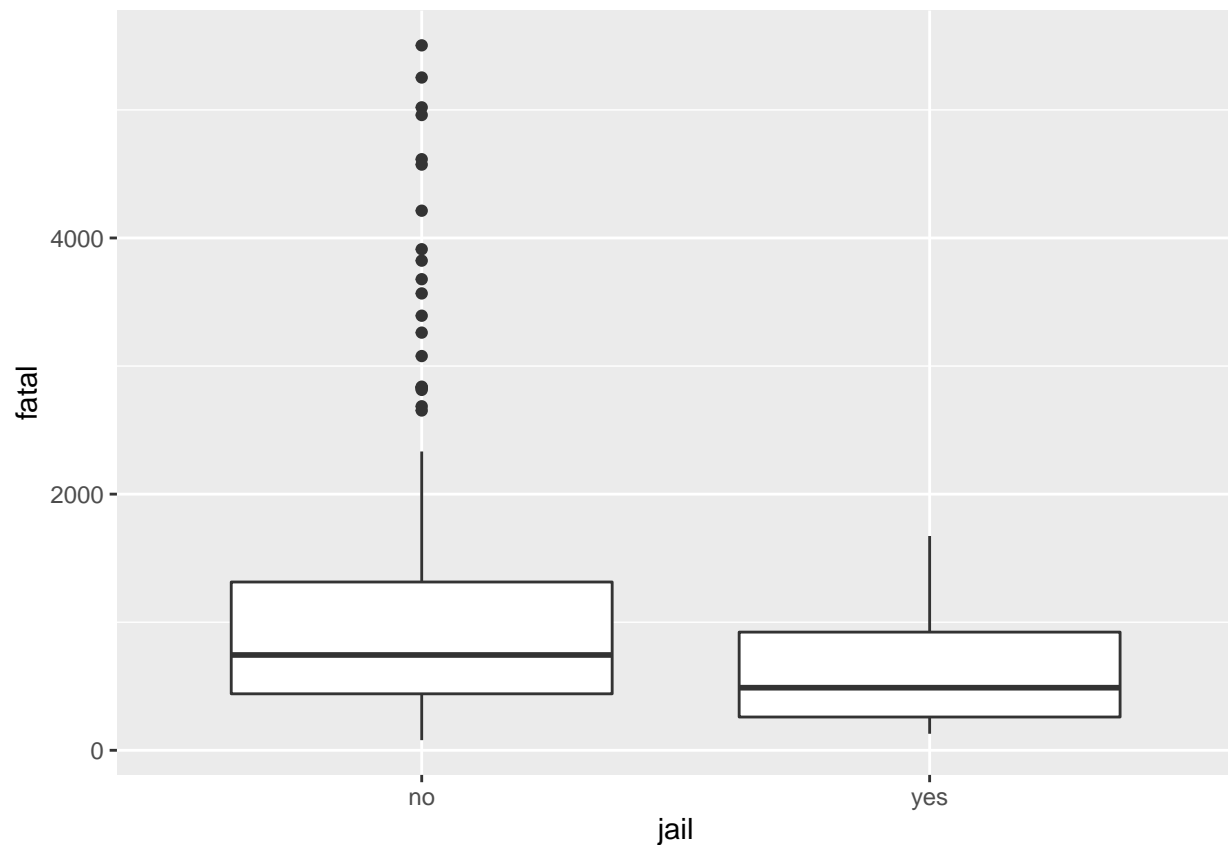
```
data['fr'] = data$fatal/data$pop*10000

ggplot(data, aes(x = jail, y = fr)) +
  geom_boxplot()
```



```
ggplot(data, aes(x = drinkage, y = fr)) +
  geom_point()
```

1. Explore this dataset and generate summary statistics (in forms of tables or plots) that you find crucial for your own interest, or for convincing the policymakers.

2. Consider only the full dataset from 1982 to 1988, propose a regression model to study whether having a mandatory jail sentence is associated with reduced traffic fatalities. In particular, you need to

- specify your model,
- state the assumptions required,
- fit the model with appropriate methods,
- conduct model diagnostics and/or sensitivity analysis,
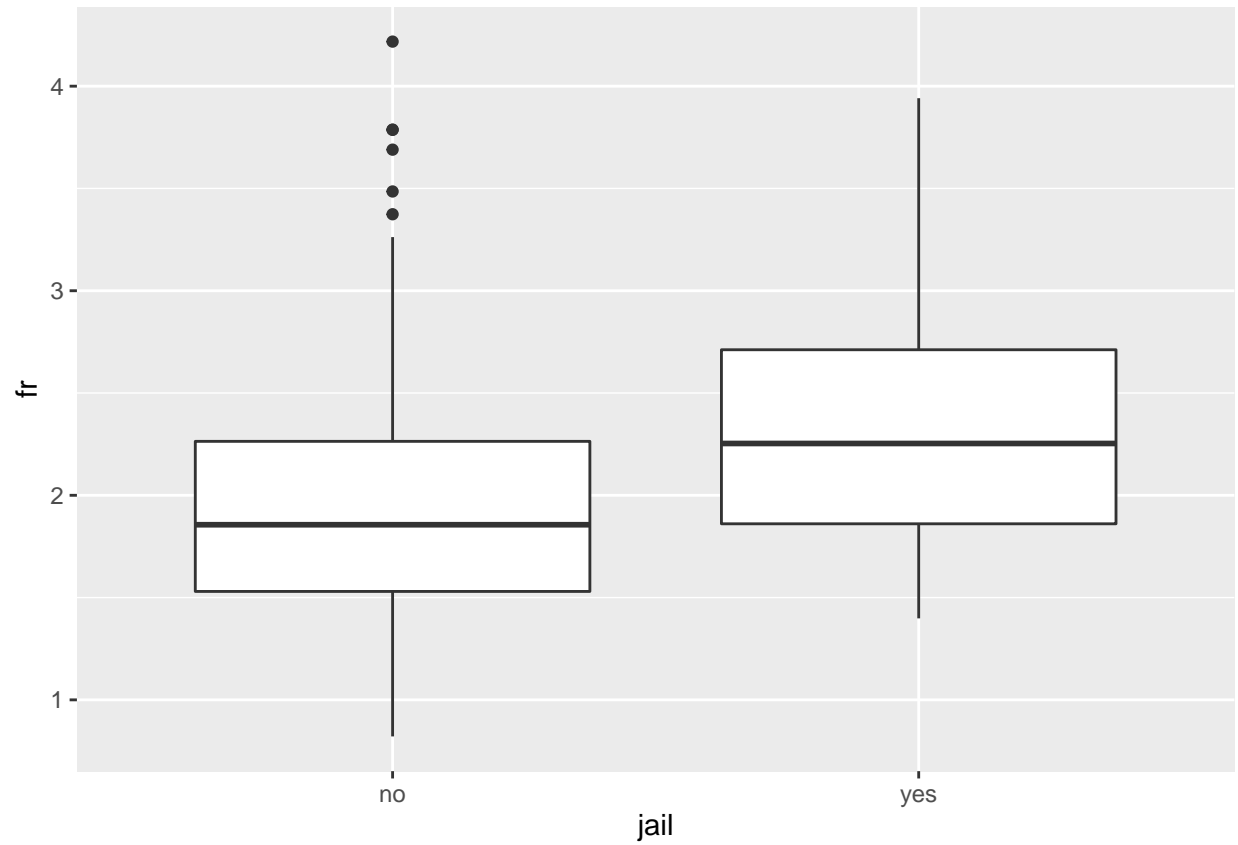- and discuss causal interpretation of the proposed models.

3. Conclude your analysis results. You may want to test a hypothesis, construct a confidence interval, or draw a confidence band.

4. Explain the implications of your results to policymakers who know little about statistics. Make suggestions if you want to.