Can mandatory jail law reduce the fatality

Team Id:11

Xi Yang (Introduction, Analysis plan, Comparison)

Shan Gao (Model fitting,F-test,Causal Statement)

Yuxiang Lin (Descriptive Analysis and Conclusion)

Ziyi Zhou (Model Diagnositic, Report)

Github Repo: https://github.com/zhouzyzyzyzy/STAT207.git

1 Introduction

1.1 Background

The global epidemic of road crash fatalities and disabilities is gradually being recognized as a major public health concern. There are approximately 40,000 highway traffic fatalities each year in the United States and an additional 2.35 million are injured or disabled, leading to the road safety a shared responsibility. In this report, we aim to study the influential factors for traffic fatalities and offer policymakers feasible suggestions in reducing fatalities.

The dataset "Fatalities" contains the traffic fatalities data for 48 U.S. states(excluding Alaska and Hawaii) from 1982 to 1988. We are interested in alcohol factors, social economical conditions and legal terms associated with vehicle crashes. Based on the literature review and dataset information, we will use fixed effect model with selected factors to study whether having a mandatory jail sentence is associated with reduced traffic fatalities.

2 Analysis Plan

2.1 Model selection

In our particular case, we are looking to observe the effects of a series of variables, based on which, we want to make suggestions to policymakers to take certain measures. We sampled all the states across the U.S. We also hope that our model would improve the national highway traffic safety within the U.S. In this case, the target population is the same with sample population. We are interested in making conclusions about whether jail sentences and other factors impact the fatality rate, then the variables would be fitted as fixed effects. The data set we used in this report is a panel data so we choose the fixed effect regression to build our model.

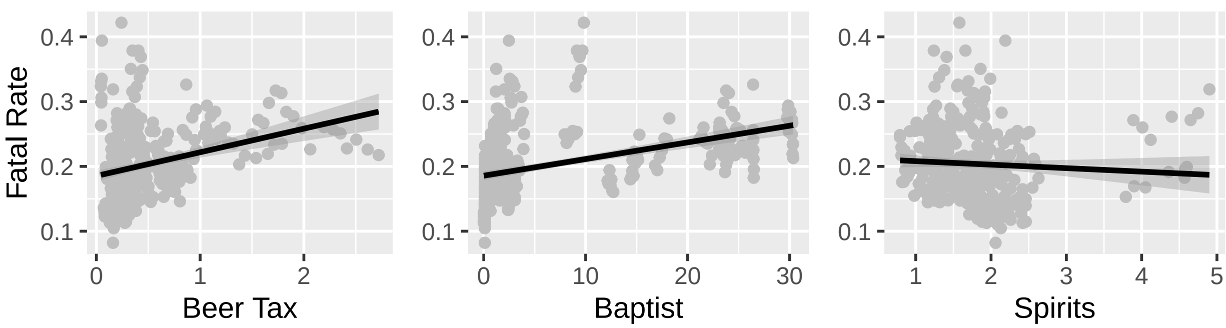
2.2 Descriptive Analysis

The original dataset has 336 observations obtained from 48 U.S. states and 7 consecutive years from 1982 to 1988. We conduct a both-direction stepwise regression to check which variables are significant enough to be involved in the model. Based on the result, we select variables state, year, jail sentence, unemployment rate, spirits consumptions, income, beer tax, average miles per driver and baptist to build out model.

We summarize three informative aspects associated with the traffic fatal rate below:

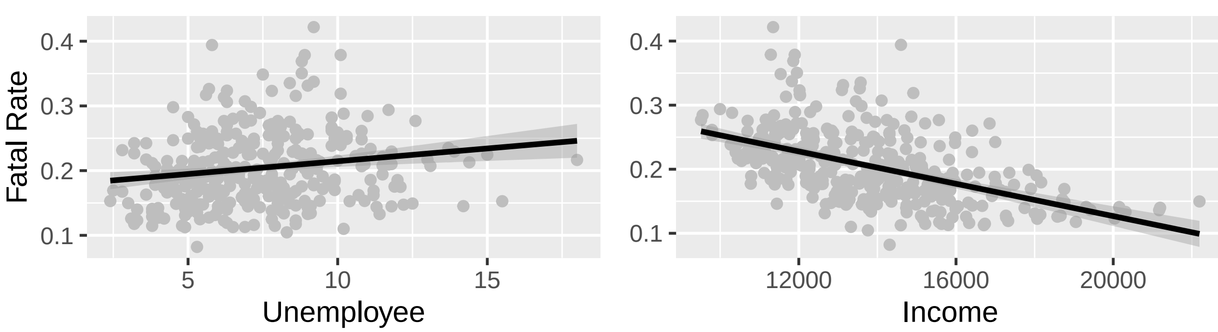
- **Alcohol factors:** Alcohol drinking is connected with drunk driving and almost half of drivers and more than 40% of passengers killed in vehicle crashes have been drinking [1]. Beer tax, spirits consumption and percentage of Baptist are indicators of alcohol consumption within a state. Beer tax and spirits consumption directly influences the beer purchase and Baptist(Southern Baptist holds a general rule of having opposed drinking alcoholic beverages). By exploring the data we find a high positive relations between Beer tax, Baptist percentage and fatal rate as well as a negative relations between spirits consumption and fatal rate `Figure 1`, indicating the necessity of considering these three factors in our model.

Figure 1

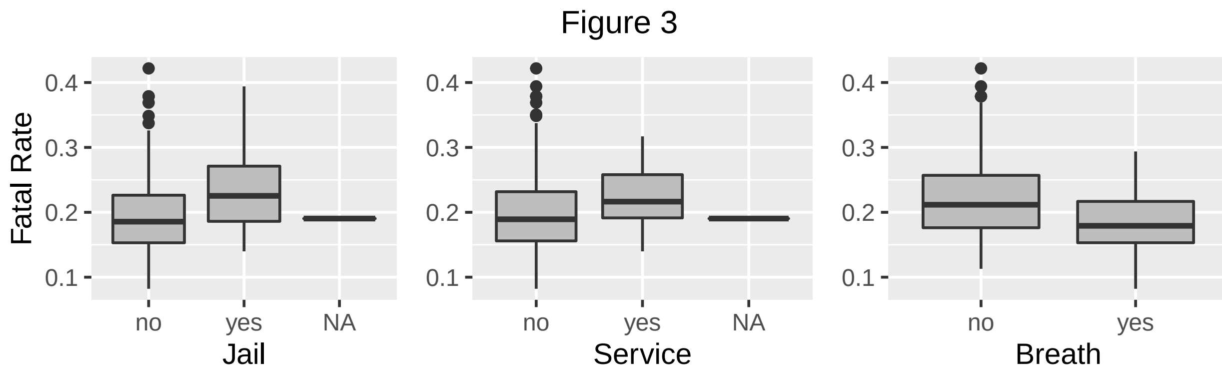


- **Economical conditions:** Early evidence suggests that the great recession and its high unemployment rates substantially reduced fatalities involving commercial vehicles [2]. When the economy was in decline there were fewer other vehicles on the road, and therefore fewer opportunities for crashes. However, an improving economy is associated with more fatal crashes. The high correlation drawn out from the dataset `Figure 2` also tells us to incorporate these two factors in our model.

Figure 2

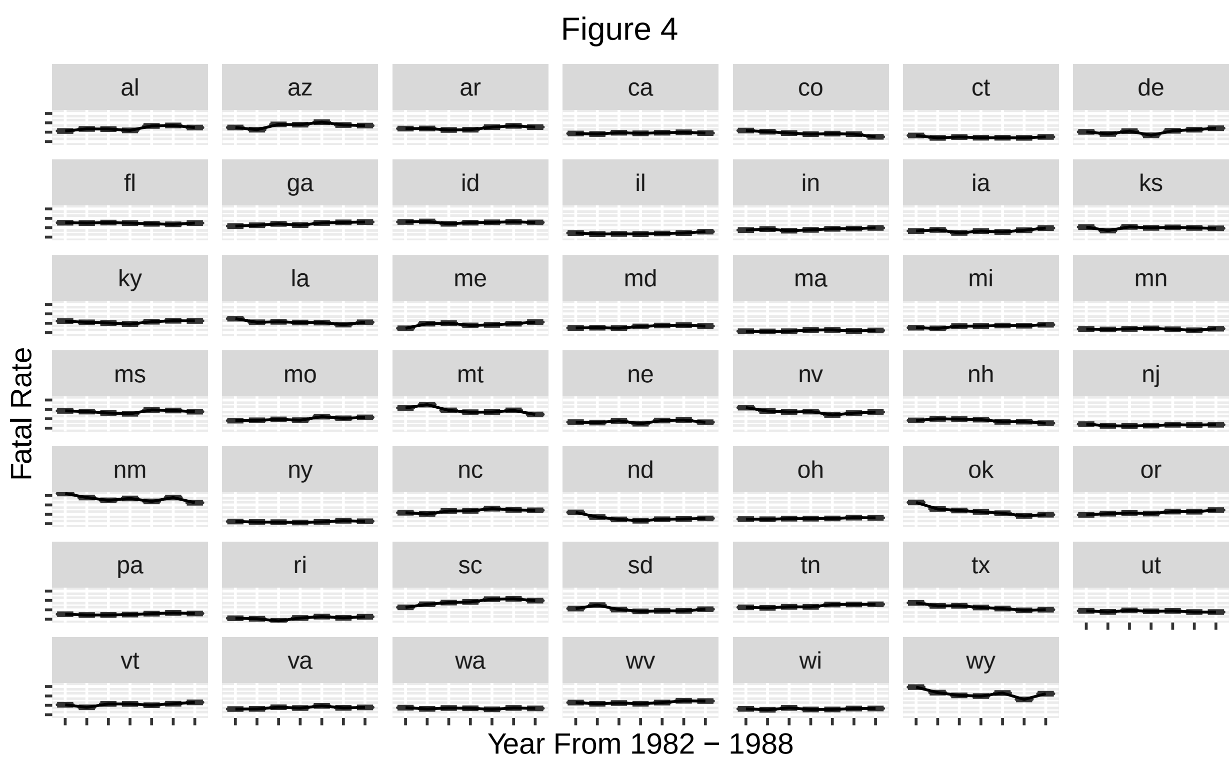


- **Legal terms:** Time served in jail is considered as a proper punishment for drunk driving and currently in all states, first-offense driving under the influence(DUI) is classified as a misdemeanor, and punishable by up to six months in jail. Despite the brevity of substituting community service, these laws have been strongly supported as providing the severity necessary for effective deterrence of drunk driving [3]. The grouped boxplots show there might exist the treatment effect under different legal penalties towards drunk driving. In addition, the preliminary breath test law affects the procedure of determining drunk driving cases and there might exist treatment effects as well.



The traffic fatalities are associated with other factors including those unobserved in our dataset. As we see from the boxplots `Figure 4`, the median of fatal rate within some states vary to some extent, it might result from the changing policy(Connecticut state revised the law of mandatory jail in 1985) as well as the unobserved variables fluctuate over time. To eliminate the source of omitted variable bias, we need to consider sources vary across states but are constant over time(e.g. In-state public attitude towards drinking) and sources vary over time but are constant across state(e.g. Manufacturing skills in car industry)

We incorporate these two factors as the fixed effects in our model to control for unobserved variables and estimation bias.



2.3 Fixed effect Model

The `fixed effects regression model`[4] is:

$$Y_{it} = \gamma X_{it} + \eta Z_{it} + \alpha_i + \beta_t + \epsilon_{it}$$

Explanation of the notation

- The index i denotes factor level of state and the index t denotes factor level of year.

- Y_{it} denotes the observed outcome of the fatality rate in the ith state and tth year, X_{it} denotes the jail variable in the ith state and tth year, Z_{it} vector denotes the other observed variables in the ith state and tth year.
- γ and η vector denotes the coefficient of jail and other interst variables.
- The α_i and β_t are entity-specific intercepts that capture heterogeneities across entities and time.
- ϵ_{it} denotes random errors. These are unobserved random variables.

Assumption: The model errors are assumed to be identically and independently distributed from a normal distribution with zero mean and equal variance.

3 Results and Discussian

3.1 Fixed Effect Model

To do the hypothesis test, we should specify the test rules of fixed effect model here:

$$H_{0j} : \beta_j = 0, j = 1, 2, \ldots, p \ v.s \ H_{1j} : \beta_j \neq 0, j = 1, 2, \ldots, p$$

and use F statistic to do the test. Table1 shows our model fitting result:

Table 1: Model Result

	Jail(Yes)	unemp	log(income)	beertax	baptist	spirits	service(Yes)	miles	breath(Yes)
Coeff	4.3e-3	-4.5e-3	1.8e-1	-3.2e-2	-3.4e-3	7.8e-2	-6.7e-4	1.2e-6	-5.9e-3
Pr(<F)	0.70	<0.001***	<0.001***	0.06	0.48	<0.001***	0.96	0.14	0.53

Table 2: Confidence interval

	Jail	unemp	log(income)	beertax	baptist	spirits	service	miles	breath
2.5%	-1.7e-2	-6.7e-2	1.1e-1	-6.5e-2	-1.3e-2	5.7e-2	-2.6e-2	-4.0e-7	-5.9e-3
97.5%	2.6e-2	-2.4e-3	2.5e-1	5.8e-4	6.1e-3	1.0e-1	2.4e-2	2.8e-6	1.1e-2

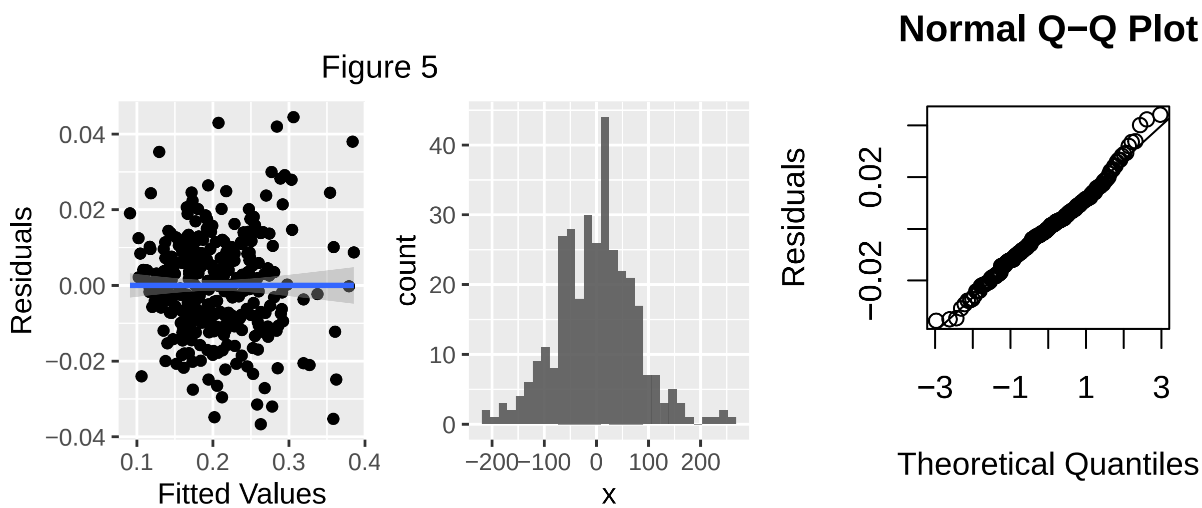
Table 2 shows the confidence interval. The estimator of Spirit consumption suggests that there is a positive correlation between traffic fatalities rate and alcohol consumption per capita. The result shows that per liter increase in alcohol consumption, the fatality rate increases 0.08 units, significant ($P<0.05$). When the consumption of alcohol goes up, we assume that the proportion of drunk drivers will also increase and, accordingly, the fatality will rise.

The income per capita has a positive correlation with the response variable. The unemployment rate has an inverse correlation with the response variable. Both are significant ($P<0.05$). The jail is not significant and has a positive correlation with the fatalities rate. The estimator of the beer tax has a negative correlation with the fatalities rate. The sign of beer tax indicates that higher the beer tax is, lower the consumption of alcohol is, and lower the fatalities rate is. However, the beer tax is not significant ($P>0.05$). The estimator of baptist and service are not significant at the 0.05 level.

3.2 Model Diagnostics

- **Independence:** From the background of the dataset, we believe that the outcomes are not independent.
- **Normality:** The right histogram in `Figure 5` depicts that the distribution is not normal and the Q-Q plot gives the same conclusion. The normality assumption is not satisfied.
- **Constant variance:** As shown in the left scatter plot in `Figure 5`, the residuals spread around the 0, but the extent of the points scattered are not equal. The assumption of constant variance is not satisfied.

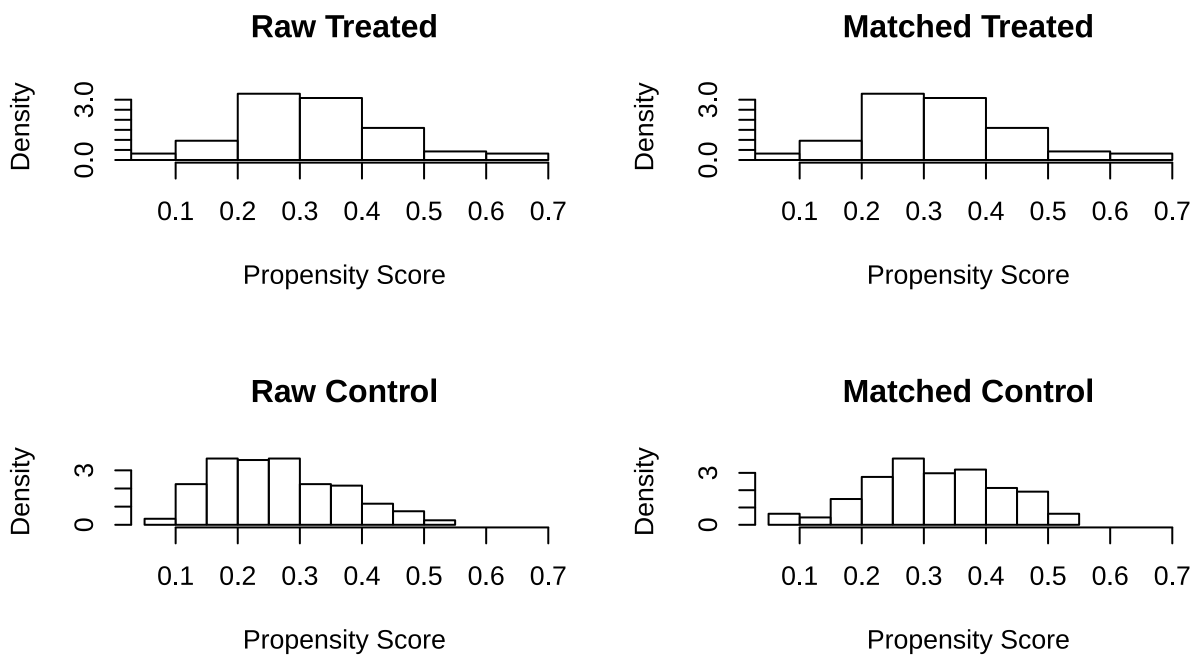
Based on those results, we can conclude that the robustness and reliability of this model need to improvement, and the reliability of conclusion above is not strong. We need to improve our model. In this case, even though jail is not significant (P>0.05), we still want to do causal inference to make sure whether it is the causal effect of the fatality rate.



3.3 Causal Inference

This experiment is designed as an observational study. The experiment units were not randomly generated nor independent with each other. Propensity score matching is a powerful technology that allows for causal inference in this type of case. In the empirical example, we adjusted for confounding using nearest neighbor matching.

Figure 6



- Estimate Propensity Scores and Adjust for Confounding
  First, we create an artificial data set that contains the following set of covariates (states which change their jail sentence policy halfway, spirits, unemployment, income, and percent of baptist) along with a treatment indicator, indicating whether or not a jail sentence was executed.
  Overlapping distributions means that there are individuals in the treatment group that are similar to those in the control group on the potential confounders. When distributions do not overlap, causal inferences are not warranted because the necessary extrapolation results in unstable estimates of effects. We examined overlap of the propensity score distributions for jail sentence to determine if an attempt at causal inference could be justified in this study. The histogram `figure 6` shows that for nearly all individuals in the control group, an individual in the jail group has a similar propensity score, and vice versa.
  Then, we select the ratio, which we set to one indicating that each person in the jail-yes group will be matched with one person in the control group (jail-no). Sample sizes `Table 3` shows the result of the procedure.

- Assess Balance
  From the percent balance improvement, which provides percentage improvement by using the matched data relative to all the data, we can see that a match is needed. The plots show the effectiveness of the matching procedure. The histogram `Figure 6` evaluates how much better the matching procedures matched the data.

Table 3: Sample sizes

Sample Sizes	Control	Treated
All	241	94
Matched	94	94
Unmatched	147	0
Discarded	0	0

- Estimate the Propensity Score-Adjusted Treatment Effects
  Finally, we use the Zelig function to create the model to evaluate the impact of being sentenced in a jail with the matched data. We can see that the jail variable is significant under 5% significance level. Hence, a causal statement can be made that the estimates of the causal effect suggest that jail sentence is expected to produce a higher fatality rate.

4 Conclusion and Suggestion

Using a large panel data of traffic fatalities with 48 states in the U.S. in 7 consecutive years, from 1982 to 1988, we investigate whether the drinking factor, the economic factor and the legal factor influence the fatality rate with fixed effect model and whether the mandatory jail law is the cause of fatality rate. Based on the result, we find that economic factors (income and unemployment) and Spirit consumption are associated with the fatality rate. We also do the causal inference and find that mandatory jail law is the causal effect of the fatality rate. The coefficient of mandatory jail law is positive, Therefore, we can make suggestions to policymakers that cancel the mandatory jail law and pay more attention to other factors like road infrastructure and economic development.

5 Reference

[1]. Ruhm, Christopher J. “Alcohol policies and highway vehicle fatalities.” Journal of health economics 15.4 (1996): 435-454.
[2]. https://www.sciencedaily.com/releases/2016/01/160121132534.htm
[3]. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=6657&context=jclc
[4]. https://www.econometrics-with-r.org/rwpd.html#drunk-driving-laws-and-traffic-deaths
[5]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4098642/

6 Appendix

```
##
## Call:
## matchit(formula = jail ~ spirits + unemp + income + baptist,
##     data = mydata, method = "nearest", ratio = 1)
##
## Summary of balance for all data:
##          Means Treated Means Control SD Control Mean Diff  eQQ Med eQQ Mean
## distance        0.3266        0.2626     0.0997    0.0640   0.0674   0.0659
## spirits         1.7004        1.7746     0.6386   -0.0742   0.1100   0.1780
## unemp           7.9383        7.1245     2.4144    0.8138   0.7000   0.8223
## income      13326.7360    14078.7541  2229.7252 -752.0180 945.8691 902.4125
## baptist         5.8978        7.6703    10.2243   -1.7724   0.6894   2.4076
##            eQQ Max
## distance    0.1704
## spirits     1.4900
## unemp       2.5000
## income   1770.9473
## baptist    13.7640
##
##
## Summary of balance for matched data:
```

```
##          Means Treated Means Control SD Control Mean Diff  eQQ Med eQQ Mean
## distance        0.3266        0.3179     0.1051    0.0088   0.0010   0.0098
## spirits         1.7004        1.7509     0.6741   -0.0504   0.0800   0.1464
## unemp           7.9383        7.5702     2.6219    0.3681   0.3000   0.3766
## income      13326.7360    13481.5622  2156.8600 -154.8262 223.6265 265.8856
## baptist         5.8978        5.2934     8.8158    0.6045   0.9407   1.3373
##            eQQ Max
## distance    0.1704
## spirits     1.4900
## unemp       2.5000
## income   1025.0234
## baptist     5.3209
##
## Percent Balance Improvement:
##          Mean Diff.  eQQ Med eQQ Mean eQQ Max
## distance    86.2887  98.5103  85.1634  0.0000
## spirits     32.0230  27.2727  17.7525  0.0000
## unemp       54.7705  57.1428  54.2044  0.0000
## income      79.4119  76.3576  70.5361 42.1200
## baptist     65.8971 -36.4563  44.4527 61.3419
##
## Sample sizes:
##           Control Treated
## All           241      94
## Matched        94      94
## Unmatched     147       0
## Discarded       0       0

## How to cite this model in Zelig:
##   R Core Team. 2007.
##   ls: Least Squares Regression for Continuous Dependent Variables
##   in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau,
##   "Zelig: Everyone's Statistical Software," http://zeligproject.org/

## Model:
##
## Call:
## z5$zelig(formula = fatal_rate ~ jail + spirits + unemp + income +
##     baptist, data = m.outCSV)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.112114 -0.031832 -0.003939  0.019534  0.200535
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.611e-01  3.260e-02  11.076  < 2e-16
## jail         3.313e-02  6.731e-03   4.922 1.91e-06
## spirits      2.579e-02  5.012e-03   5.146 6.86e-07
## unemp       -4.103e-03  1.432e-03  -2.865 0.004661
## income      -1.386e-05  2.004e-06  -6.917 7.58e-11
## baptist      1.473e-03  4.247e-04   3.467 0.000656
##
## Residual standard error: 0.04601 on 182 degrees of freedom
```

```
## Multiple R-squared:  0.3833, Adjusted R-squared:  0.3663
## F-statistic: 22.62 on 5 and 182 DF,  p-value: < 2.2e-16
##
## Next step: Use 'setx' method
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = fatal_rate ~ jail + state + year + unemp + log(income) +
##     beertax + +baptist + spirits + service + miles + breath,
##     data = Fatalities, effect = "twoways", model = "within",
##     index = c("state", "year"))
##
## Unbalanced Panel: n = 48, T = 6-7, N = 335
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.03559012 -0.00819242  0.00021492  0.00742848  0.04416078
##
## Coefficients:
##               Estimate  Std. Error t-value  Pr(>|t|)
## jailyes      4.3060e-03  1.1083e-02  0.3885   0.69794
## unemp       -4.5451e-03  1.0868e-03 -4.1820 3.900e-05 ***
## log(income)  1.8083e-01  3.4039e-02  5.3125 2.253e-07 ***
## beertax     -3.2026e-02  1.6636e-02 -1.9251   0.05526 .
## baptist     -3.4215e-03  4.8616e-03 -0.7038   0.48216
## spirits      7.8459e-02  1.1091e-02  7.0739 1.276e-11 ***
## serviceyes  -6.6688e-04  1.2707e-02 -0.0525   0.95818
## miles        1.1867e-06  8.0732e-07  1.4699   0.14275
## breathyes    2.7845e-03  4.4411e-03  0.6270   0.53120
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.10289
## Residual Sum of Squares: 0.055716
## R-Squared:      0.45851
## Adj. R-Squared: 0.33508
## F-statistic: 25.5905 on 9 and 272 DF, p-value: < 2.22e-16
```