Effects of Medical Marijuana Legalization of Traffic Fatalities Using Bayesian Framework

Nhat Hoang Pham

University of Colorado, Denver

MATH 7397

Professor Joshua French

Abstract

Over the last few years, marijuana has become legal for recreational use in 18 states. However, policy makers and researchers still have little known about its effect since identifying the effect for a recent phenomenon requires more data after legalization. In this paper, mainly using FARS data and Bayesian framework , I examine the relationship between medical marijuana legalizations (hereafter MMLs) and traffic fatalities. I find that the traffic fatality rate decreased by about 9% fatalities per 100,000 people after states adopted medical marijuana laws.

Keywords: medical marijuana legalization, traffic fatalities, Bayesian Statistics.

Effects of Medical Marijuana Legalization of Traffic Fatalities Using Bayesian Framework

Introduction

At the federal level, on July 14th, 2021, the proposed Cannabis Administration and Opportunity Act would remove marijuana from Controlled Substances Act and begin regulating and taxing it (Fandos 2021). Public opinion polls show that roughly 70 % of Americans supporting the legalization of marijuana, this number has doubled from 2000s (Gallup 2020).

Opponents argue that marijuana poses a substantial risk to individual health, and public health as well. At the individual level, using and smoking marijuana daily increases chronic cough and phlegm frequent chronic bronchitis episod (Division, 2017). On the public health side, opponents also have expressed concern that MMLs can increase the use of marijuana among teenagers, which can lead to negative and long-lasting effect on their cognitive development (CDC). Some opponents are simply morally against the legalization of marijuana.

On the flip side, proponents of legalizing marijuana say that the evidence proving the adverse health effect of marijuana is light-to-moderate (National Academies of Sciences, Engineering and Medicine 2017). Furthermore, in comparison with harmful effects from alcohol and opioid and considering substitutability of marijuana, those above-mentioned negative health effects are smaller, and by increasing the use of marijuana, we can have some public health gains. Proponents also point out that the annual cost of enforcing marijuana ban is up to hundreds of billions of dollars (American Civil Liberties Union 2019). In addition, for whom are convicted of marijuana-related crime, their job outlook is poor (Pager 2003; Agan and Starr 2018; Dobbie et al. 2018; Mueller- Smith and Schnepel 2021; Agan, Doleac and Harvey 2021).

Studies conducted in laboratory settings have demonstrated that the use of cannabis adversely affects driving-related skills such as distance perception, reaction time, and hand-eye coordination (Kelly et al. 2004; Sewell et al. 2009). However, studies conducted using driving simulators or on actual driving courses have yielded inconsistent evidence regarding whether these impairments in driving-related skills result in a higher risk of collisions (Kelly et al. 2004; Sewell et al. 2009). Individuals driving under the influence of marijuana tend to lower their speed, avoid hazardous maneuvers, and increase their distance from other vehicles, indicating the use of compensatory measures (Kelly et al. 2004; Sewell et al. 2009).

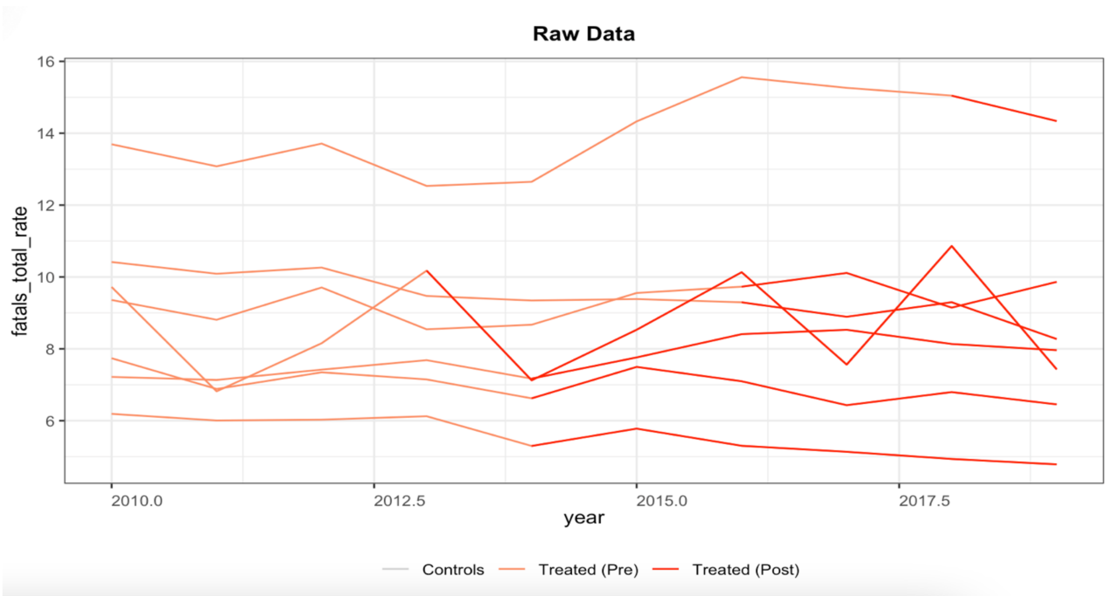
The reasons for society to adopt MMLs depends on whether the benefits of MMLs outweigh any external costs. In this study, I examine the impacts of MMLs on traffic fatalities. Using state-level data from Fatality Analysis Reporting System in the period 2010-2019. I find that the traffic fatality rate decreased by about 9% trafficfatalities rate people after states adopted medical marijuana laws. This result preliminarily suggests that marijuana can reduce the traffic fatalities though we are unsure about its mechanism. One possible answer is that marijuana is a substation to other drugs, especially alcohol. By reducing alcohol consumption, and less people driving under alcohol influence were on the street, then it would lower traffic fatalities rate at the state level.

Exploratory Data Analysis

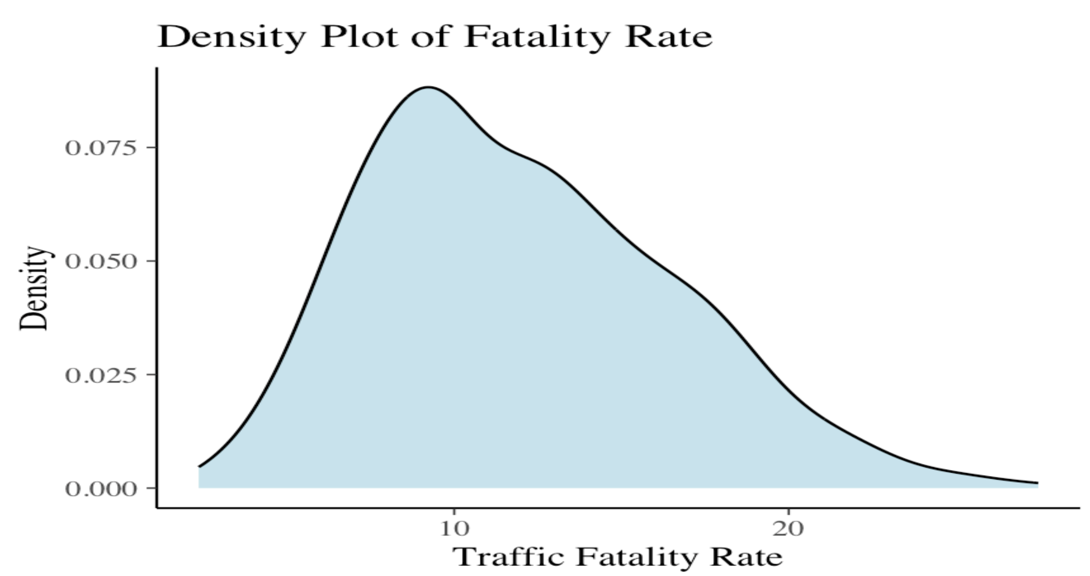
To analyze the data, I began by choosing total traffic fatalities rate as my response variable and the following predictors. There are 510 observations(each state has ten years) and 12 variables in total.

	Mean	*SD*	*Description*
Response variable *Total Fatalities*	11.97	4.57	Fatalities per 100,000 people
Predictors			
MMLs	0.47	0.5	Equals one if a state had a medical marijuana law in a given year and zero otherwise
Year	Range: 2010-2019	//	Calendar year
Driver License	4.28e+06	4.57e+06	Number of driver license in a state
Total Miles	6.05e+04	6.27e+04	Total vehicle miles in a state
Income	47.18	5.97	Median income in a state (thousands of dollars)
Age	38.06	2.41	Median age in a state
Population	6.26e+06	7.06e+06	Total population of a state
Unemployment Rate	5.788	2.25	State unemployment Rate
Decriminalized	0.3	0.46	Equals one if a state had a marijuana decriminalization law in a given year and zero otherwise
Texting ban	0.74	0.432	Equals one if a state had a cell phone texting ban in a given year and zero otherwise
Miles Driven	14.5	2.78	Vehicle miles driven per licensed driver (thousands of miles)

First, I created graph to see the raw trend of traffic fatalities rate for some states adopting MMLs in the period(2010-2019), it is hard to see any obvious trends in traffic fatalities after states enacted MMLs.



Since Bayesian framework needs some assumption to be made at the begining, a quick look



at density plot helps me decide which assumptions I should be making in the analysis.

The density plot is a bit left-skewed normal-shaped, so initially, my first assumption is that the data distribution is normal without any modifications. This initial plan changed later for better posterior check result.

Collinearity Assessment

First, I used GGpair function to have a bigger picture of our data. As we can see below, there are a few high correlation (over 0.95) that are noticeable between population, vehicles miles, and number of driver license variables. That is understandable since they are very closely related.

Secondly, I used Variable Inflation Indicator (VIF) to determine which variables I should drop. The VIFs for Driver License, Population and Total Miles is dramatically high,(159.86 , 59.89, 118.36 respectively) way over 5. The VIF suggests that Driver License, Population and Total Miles variables should be dropped.

Predictors Variables	***VIF***
MMLs	1.672
Year	3.249
Driver License	159.858
Total Miles	59.89
Income	1.525
Age	1.207
Population	118.358
Unemployment Rate	3.196
Decriminalized	1.13
Texting ban	1.265
Miles Driven	2.038

Since we have Miles Driven as a very good predictor for driving behavior and density of population in states, so I decided to drop *Driver License, Population and Total Miles* variables.

Variable Selection

Using BVS function from libray BayesVarSec, we have a table of inclusion propabilities for choosing predictors. As we can see below, there are two variables with lowest probabilities I should consider to exclude: *Texting_ban and Unemployment rate* predictors.

Inclusion Probabilities:	*Inclusion prob.*
MML_1	0.4695
year	0.7131
Income	1
age	0.738
unemployed	0.2311
decriminalized	0.9999
texting_ban	0.24
VM_per_driver	1

In the end, I chose the final model based on information criteria (WAIC, and LOOIC) from four initial models. The first model includes every variables, the second model only excludes *Unemployment Rate*, the third model only excludes *Texting_ban* and the fourth model excludes both *Unemployment Rate and Texting_ban* variables.

The fourth model has the smallest WAIC and LOOIC statistics, so we conclude that it is the best model. On the other hand, using loo_compare function, the differences of WAIC and LOOIC statistics between all four models are pretty small (less than 2), any models could be reasonable. I decided to go with the fourth model since as of 2019, there are 48 states with texting ban in the U.S., so there is little to explore the effect of texting ban on traffic fatalities.

	elpd_diff	***se_diff***
stan_glm_final_4	0.0	0.0
stan_glm_final_2	-0.5	1.3
stan_glm_final_3	-1.0	0.5
stan_glm_final_1	-1.5	1.4

After all, my initial choice for model is:

$$Fatalitie\_rate|\beta,\sigma^2 \sim N(\mu_i,\sigma^2)\ i.i.d \text{ for i = 1,2,3…n with:}$$

$$\mu_i = \beta_0 + \beta_1 MML\_1_i + \beta_2 \mathrm{year}_i + \beta_3 Income_i + \beta_4 decriminalized_i + \beta_5 MilesDriven + \beta_6 Age$$

Prior Distribution:

$$\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6 \sim N(0.0,10^2)\ i.i.d$$

And default prior for σ^2

However, after thorough posterior checking, I decided to use log-transformation for response variable: traffic fatalities rate since it provides a better fit for posterior sample.

Thus, my final model is:

$$Log\_Fatalitie\_rate|\beta,\sigma^2 \sim N(\mu_i,\sigma^2)\ i.i.d \text{ for i = 1,2,3…n with:}$$

$$\mu_i = \beta_0 + \beta_1 MML\_1_i + \beta_2 \mathrm{year}_i + \beta_3 Income_i + \beta_4 decriminalized_i + \beta_5 MilesDriven + \beta_6 Age$$

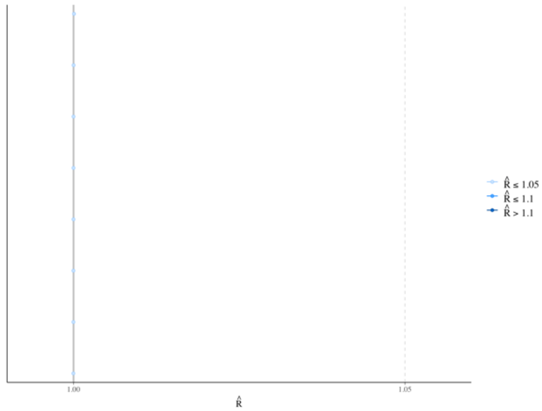
Prior Distribution:

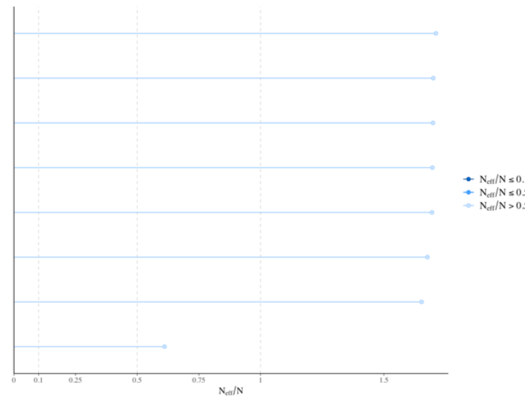
$$\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6 \sim N(0.0,10^2)\ i.i.d$$

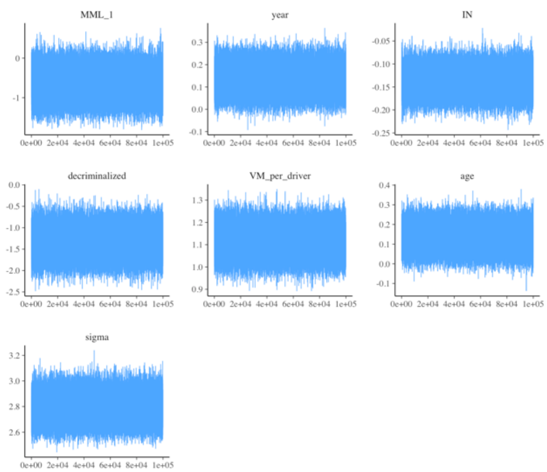
And default prior for σ^2

Convergence

One important aspect of Bayesian Analysis is to assess the convergence of Markov Chain Monte Carlo chains. Running model using 4 chains with 50,000 iterations, I used Gelman-Rubin statistics (Rhat) values, the effective sample size ratios, Heidelberg & Welch test, Raftery-Lewis test, Geweke test, trace plots in order to determine whether the MCMC chains converge or not.








```
              Stationarity start     p-value
              test         iteration
(Intercept)   passed           1     0.2398
MML_1         passed           1     0.5606
year          passed           1     0.2410
IN            passed           1     0.6881
decriminalized passed          1     0.1751
VM_per_driver passed       20001     0.0618
age           passed           1     0.6415
sigma         passed           1     0.5725

              Halfwidth Mean      Halfwidth
              test
(Intercept)   passed    -38.6980 4.10e-02
MML_1         passed     -0.0938 1.26e-04
year          passed      0.0199 2.06e-05
IN            passed     -0.0187 9.65e-06
decriminalized passed    -0.1072 1.12e-04
VM_per_driver passed      0.0914 2.43e-05
age           passed      0.0160 2.18e-05
sigma         passed      0.2369 5.81e-05
```

```
> raftery.diag(post_final)

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

                Burn-in  Total Lower bound  Dependence
                (M)      (N)   (Nmin)       factor (I)
 (Intercept)    6        8854  3746         2.36
 MML_1          6        8580  3746         2.29
 year           6        8390  3746         2.24
 IN             6        8294  3746         2.21
 decriminalized 6        8428  3746         2.25
 VM_per_driver  6        8674  3746         2.32
 age            6        8890  3746         2.37
 sigma          6        8194  3746         2.19
```

```
> geweke.diag(post_final)

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

   (Intercept)          MML_1           year             IN decriminalized  VM_per_driver            age          sigma
       -1.5889        -1.1395         1.5369        -0.4639         0.8354         1.5246         1.4489        -0.5972
```

As we can see from above, the model passed all the tests of MCMC convergence. Gelman-Rubin statistics (Rhat) isclose to 1. The effective sample size ratio is way above 0.1.

The trace plot is wiggly, The dependence factor from Rafty-Lewis is less than 5, then Z scores from Gewke are between -2 and 2. We can conclude that the MCMC chains has converged.

Assessing model fit. Structure checking, posterior predictive checks

Model checking is important in statistical analysis since a poor model fit can lead to erroneous conclusion. If my model is appropriate, then the replicated data should look like the observed data. In this section, I did some posterior predictive checks on abovementioned log-transformation model:

$$\mathbf{Log\_Fatalitie\_rate}|\boldsymbol{\beta}, \boldsymbol{\sigma}^2 \sim \mathbf{N}(\boldsymbol{\mu}_i, \boldsymbol{\sigma}^2)\ \mathbf{i.i.d}\ \ \mathbf{for\ i = 1,2,3\ldots n\ with:}$$

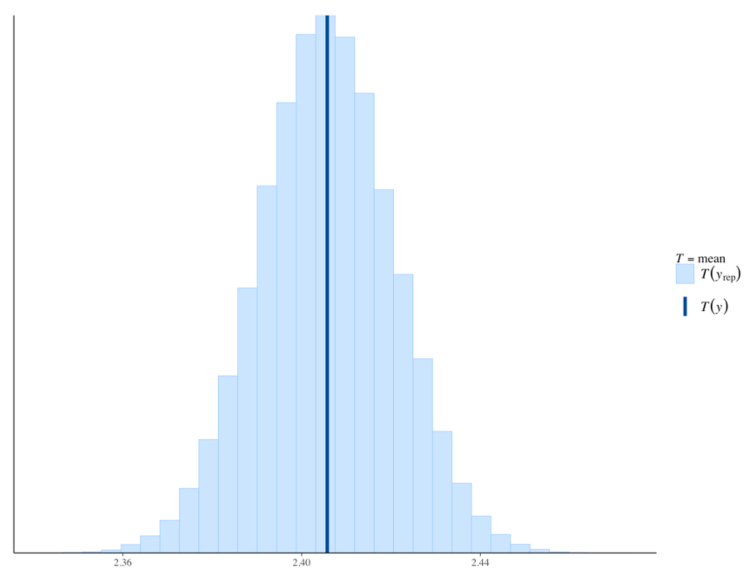
$$\boldsymbol{\mu}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\mathbf{MML\_1}_i + \boldsymbol{\beta}_2\mathbf{year}_i + \boldsymbol{\beta}_3\mathbf{Income}_i + \boldsymbol{\beta}_4\mathbf{decriminalized}_i + \boldsymbol{\beta}_5\mathbf{MilesDriven} + \boldsymbol{\beta}_6\mathbf{Age}$$

Prior Distribution:

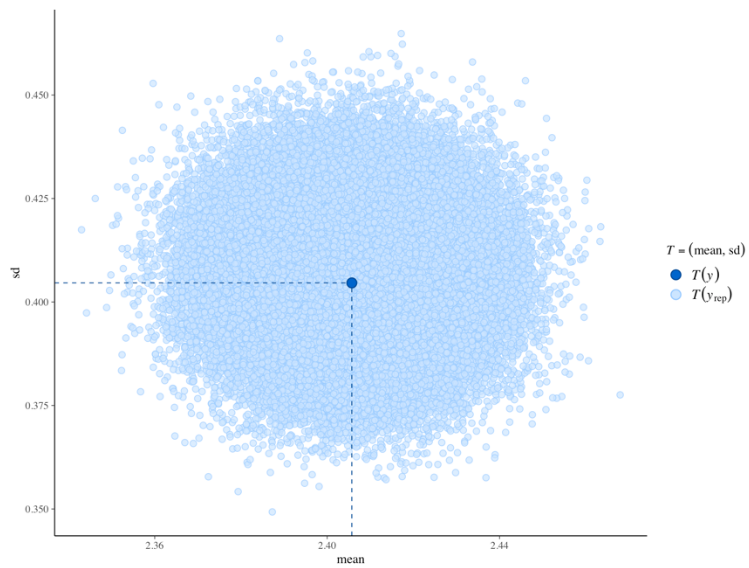
$$\boldsymbol{\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6 \sim N(0.0, 10^2)\ i.i.d}$$

And default prior for σ^2

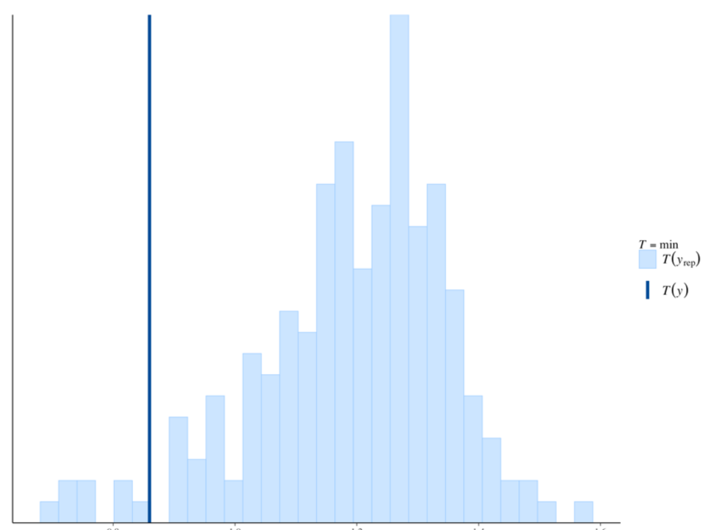
First, I used the test quantity is T(y) = mean(y), I compared T(y) for the observed data to that of 200 replicated data sets. A histogram of the discrepancy measure results is show below:



Secondly, I used the test quantity is T(y) = (mean, standard deviation). A scatter plot of the test quantity is below:

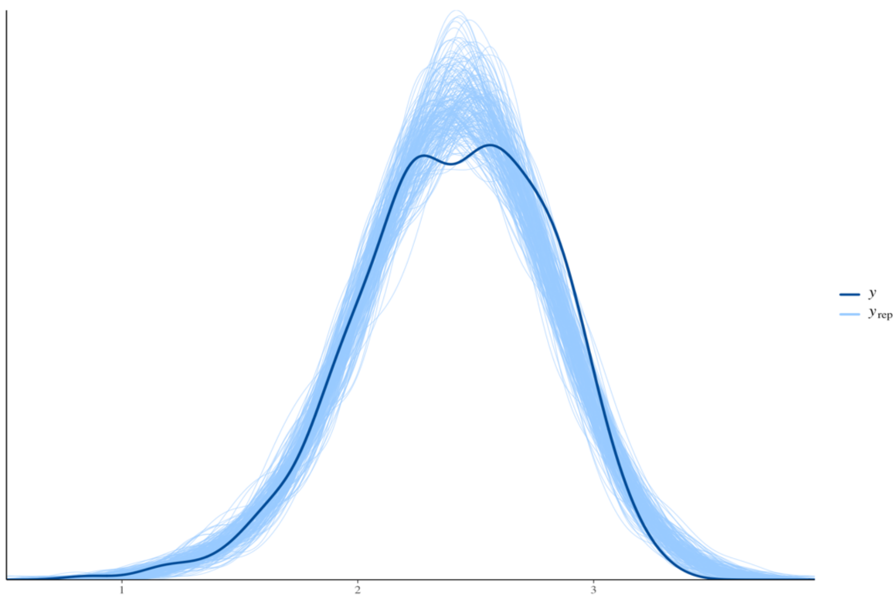


Third, I used T(y) = min (y). The p-value is 0.047, which is below 5.

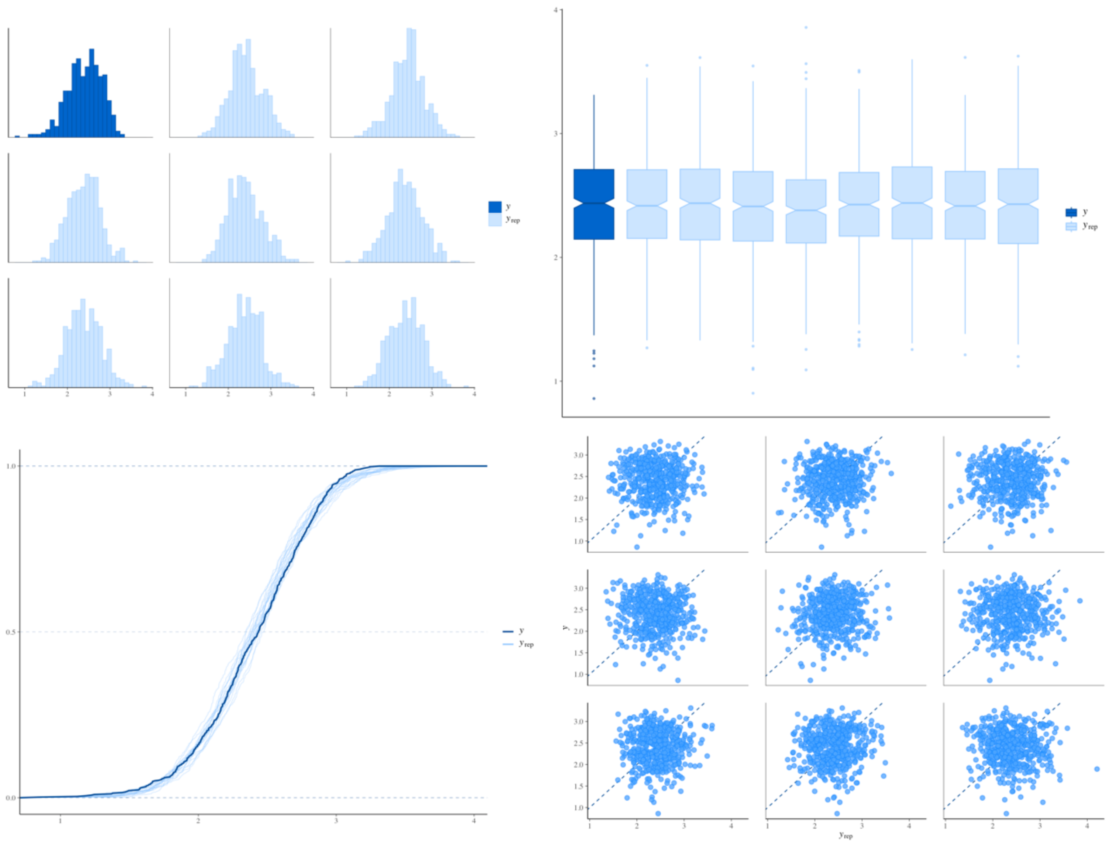


There are little discrepancy between T(y) and T(y_rep). Based on these test, the replicated data highly reflected the observed data.

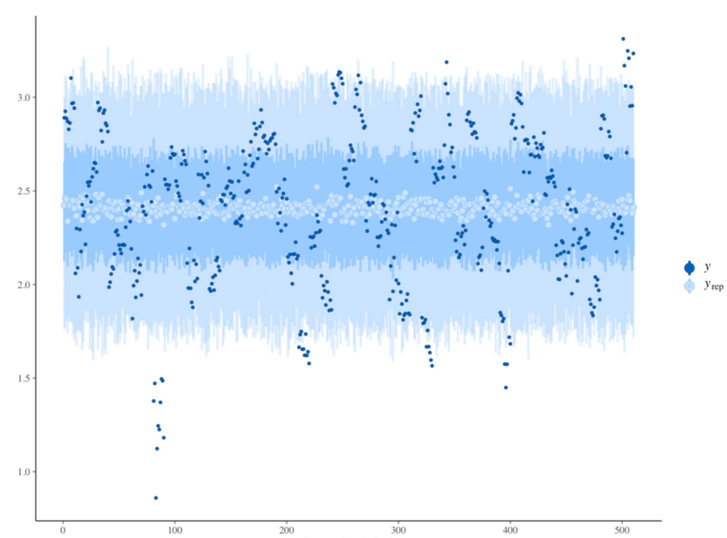
Next, we have a plot comparing the density of y and y_rep. In the plot below, the dark line is the distribution of the observed outcomes y and each of the 200 lighter lines is the kernel density estimate of one of the replications of y from the posterior predictive distribution (i.e., one of the rows in yrep). The darker line falls in between the lighter lines, and this plot makes it easy to see that this model succeeds accounting for most of the observation in y.



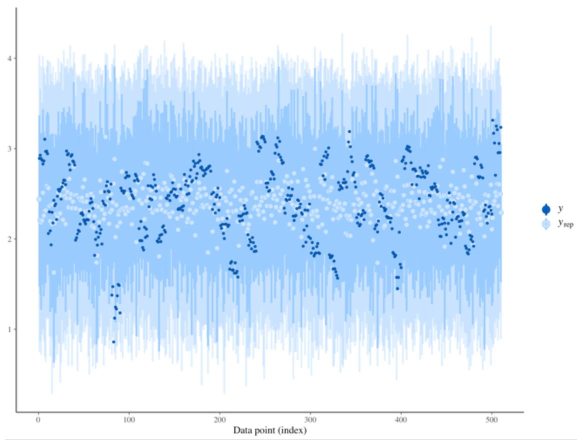
The comparisons using histogram, boxplot, the ecdf, scatter plots between y and y_rep also gives us a promising result, but we have few outliers from the low side. All plots are below:

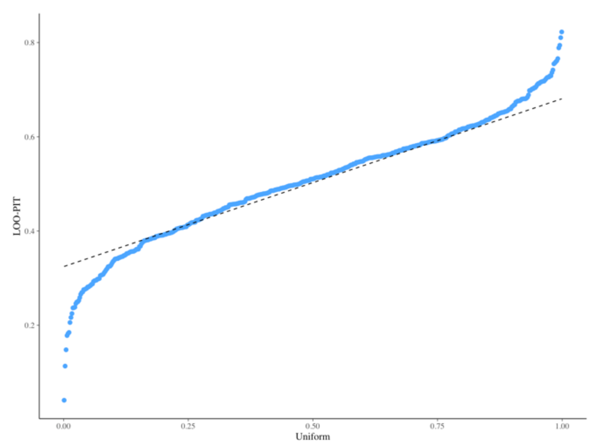


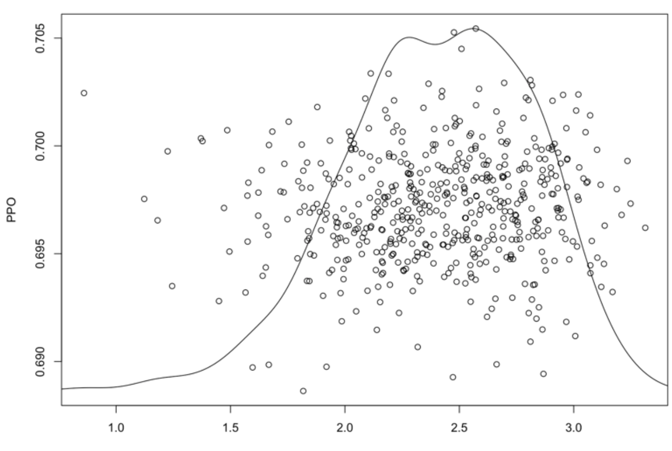
When plotting comparison of observed data and 90% predictive cpis, few low-outliers of traffic fatalities rate at state-year level are noticeable. An outlier analysis should be conducted at this point,



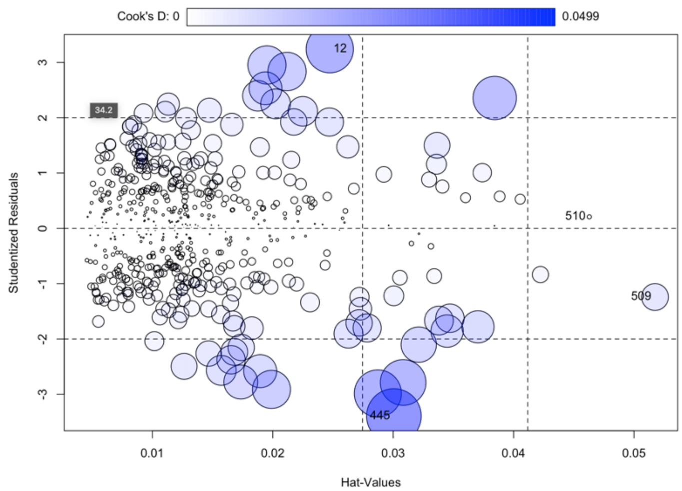
I also did some leave-one-out posterior checks. Below are graphs of leave-one-out prediction intervals, leave-one-out quantiles, and scale density to match scale of PPO.

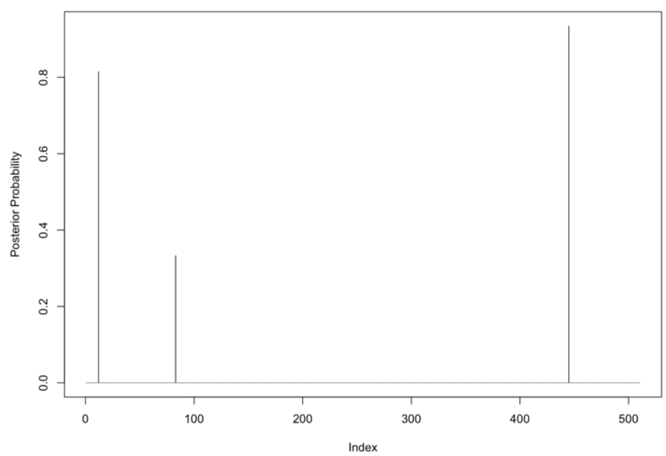






The model was checked for the effect of influential observations. First, using the influence plot, I identified there are several outliers, the observations with indices are 12 and 445. Using Bayes Outlier Detection function from BAS library really confirmed the statement.

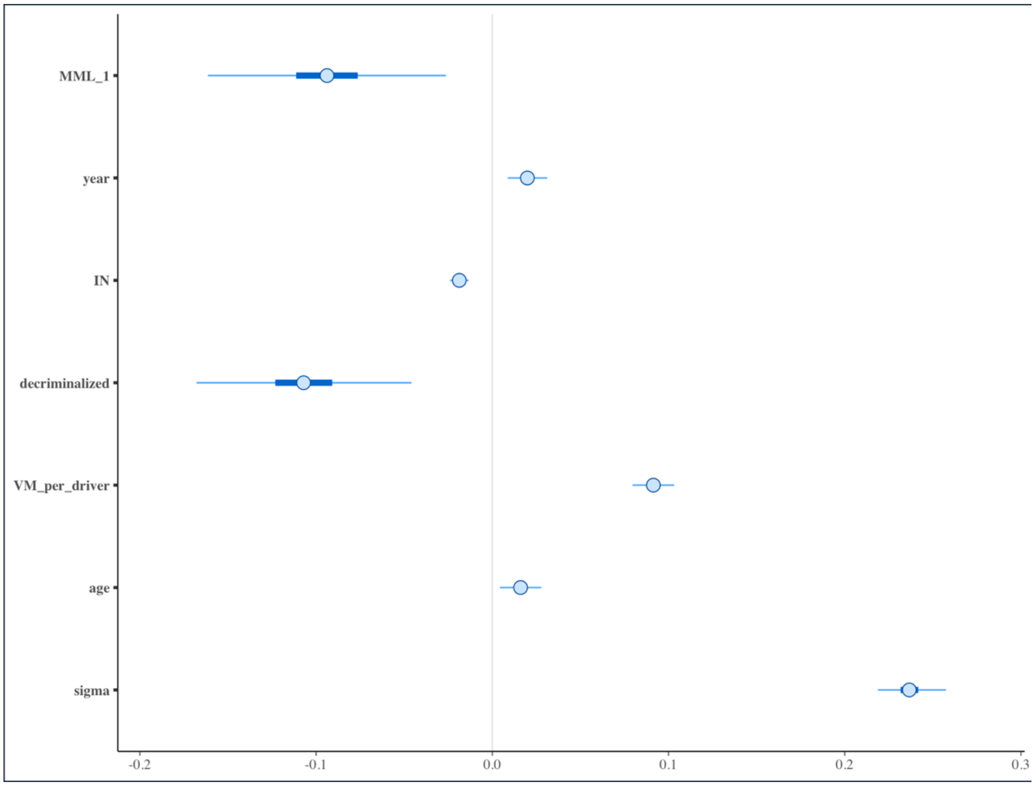




A further digging into data helps me to identify that those observation are UTAH 2014, and ALASKA 2011. There were no action taken since I believe they are still population interest of my analysis. Besides, removing those outliers doesn't dramatically change the sign, or magnitude of our coefficients of interest.

Results Interpretation, and Conclusion

Below is the graph illustrating 99% credible interval of our coefficients. All coefficients are far away from zero, so we are confident that we have some significant coefficients.



Below is table of coefficients from our final model:

	Mean	**Sd**	**2.5%**	**97.5%**
Intercept	-38.66	8.66	-55.6	-21.73
MML_1	-0.0937	0.026	-0.145	-0.04
year	0.02	0.004	0.011	0.028
IN	-0.019	0.002	-0.023	-0.015
decriminalized	-0.11	0.024	-0.15	-0.06
Miles	0.09	0.004	0.082	0.1
age	0.02	0.004	0.007	0.025
sigma	0.23	0.0074	0.22	0.25

The results suggest that the adoption of medical marijuana laws (MMLs) is associated with a decrease in traffic fatalities rate, on average, by approximately 9.37%, holding other variables constant. This finding may be due to several factors, such as reduced alcohol consumption, increased use of ride-sharing services, or reduced stress and anxiety, which could lead to more cautious driving behavior. However, it's crucial to see that this is an observational study, and causality cannot be established without further research.

On average, holding other variables constant, the finding that a one-year increase in calendar time is associated with a 2% increase in traffic fatalities rate is alarming. This suggests that there may be other factors at play that are increasing the risk of traffic fatalities over time, such as changes in driving habits, road conditions, or vehicle technology.

Holding other variables constant, The result that states with higher median incomes are associated with a decrease in traffic fatalities rate by approximately 1% on average, For example, higher-income states may have better infrastructure and safety measures in place, or residents may have greater access to public transportation, reducing the number of vehicles on the road.

The finding that a one-thousand mile increase in the number of driven miles per driver is associated with a 9% increase in traffic fatalities rate, on average, holding other variables constant, suggests that longer distances traveled are associated with a greater risk of traffic fatalities. The reasons could be fatigue, increased exposure to hazardous road conditions.

Finally, the finding that a one-year increase in the median age of a state is associated with a 2% increase in traffic fatalities rate, on average, holding other variables constant, is also worrying. This suggests that there may be age-related factors, such as declining vision or slower reaction times, that are contributing to a high risk of traffic fatalities. One weak prediction we can make would be a constraint marijuana legalization can be beneficial to our society, especially on reducing traffic fatalities rate.

In conclusion, these results highlighted the complex interplay between various factors that contribute to traffic fatalities rates, and that requires a comprehensive approach to reducing traffic fatalities is needed, including approaches such as better infrastructure, improved driving education, and increased access to public transportation. Using Bayesian framework, the effect of legalizing marijuana on traffic fatalities rate is interesting, it reduces traffic fatalities rate. This preliminary result implies that marijuana could be a substitution to alcohol. However, for future research opportunities, causality studies need to be established to for specific public recommendations by using two-way fixed effect difference in differences. Data on traffic fatalities rate with high alcohol concentration in blood is needed for future studies. Besides, studies about how legalizing marijuana at recreational level change the whole landscape of society, not only on traffic fatalities, but also on other drug behaviors, suicide rate, alcohol consumption.

References

Anderson, D. Mark, Benjamin Hansen, and Daniel I Rees. 2013 “Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption.” Journal of Law and Economics, 56(2): 333-369.

Fandos, Nicolas. 2021. “In a Milestone, Schumer Will Propose Federal Decriminalization of Mar- ijuana.” New York Times , Accessed on December 14 at https://www.nytimes.com/2021/07/14/ us/politics/marijuana-legalization-schumer.html?action=click&module=Top%20Stories& pgtype=Homepage

Vehtari, A. (n.d.). Online material. https://avehtari.github.io/