# BT2101 GA4 Group 67 Submission

Lo Zhi Hao

2022-10-23

# 1 Fixed Effect Regression using R

Please use the Fatalities dataset from the AER package in R to answer this question. Please carefully read the document of data description. Government wants to reduce traffic fatalities by increasing the tax on alcoholic beverages. Thus, they need help understanding the relationship between the traffic fatality rate and the alcohol taxes.

```
## Setting up the environment for further studies

## install.packages("wooldridge")
## install.packages("dplyr")
## install.packages("ggplot2")
## install.packages("AER")
## install.packages("plm")

library(AER)
library(dplyr)
library(knitr)
library(corrplot)
library(ggplot2)
library(plm)

## Downloading the dataset
data('Fatalities')
summary(Fatalities)
head(Fatalities)
Fatalities <- Fatalities %>% mutate(fatal_rate = fatal / pop)
```

The basic model is:

FatalityRateit = β0 + β1BeerTaxit x μit (1)

a. Suppose that the fatality rate FatalityRate is defined as the number of vehicle fatalities divided by population. Please generate FatalityRate and estimate the basic model (i.e. a model without any fixed effects). Please interpret all the coefficient estimates in your regression table

```
## creating the linear model

model <- lm(fatal_rate ~ beertax, data = Fatalities)
summary(model)
```

```
##
## Call:
## lm(formula = fatal_rate ~ beertax, data = Fatalities)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.091e-04 -3.777e-05 -9.436e-06  2.855e-05  2.276e-04
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.853e-04  4.357e-06  42.539  < 2e-16 ***
## beertax     3.646e-05  6.217e-06   5.865 1.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.437e-05 on 334 degrees of freedom
## Multiple R-squared:  0.09336,    Adjusted R-squared:  0.09065
## F-statistic: 34.39 on 1 and 334 DF,  p-value: 1.082e-08
```

The relationship is as follows:

`fatal_rate` = 1.853e-04 + 3.646e-05 x `beertax`

Multiple R squared is 0.09336, and Adjusted R squared is 0.09065.

When no tax is charged on beer, the coefficient for fatality rate is 1.853e-04, which suggests that the fatality rate due to traffic incidence in a year in USA when the total amount of beer tax is 0 is 1.853e-04. Every 1 dollar increase in beer tax per case is associated with a 3.646e-04 increase in fatality rate in a year in USA. The results for beta0 is statistically significant (p-value = 2e-16 < 0.05), which suggests we are statistically confident that it is different from 0. For beer tax, the coefficient is statistically significant (p-value = 1.08e-08 < 0.05), which suggests that we are statistically confident that the effect of beer tax on fatality rate is different from 0.

b. Determine how many years and different states there are in the data set. Is it a balanced panel? Explain. (Hints: A panel dataset is considered “balanced” when all entities are observed for all time periods. If there are any missing values, it is called an unbalanced panel.)

```
## determining number of years

years <- length(levels(Fatalities$year))

## determining number of states

states <- length(levels(Fatalities$state))

## creating a dataframe to record the data

data1982 <- Fatalities %>% filter(year == 1982)
bool <- is.pbalanced(Fatalities)

df <- data.frame(years, states, bool, length(data1982$state))
names(df) <- c("years", "states", "Balanced Or Not", "total number of observations per year")
kable(df)
```

| years | states | Balanced Or Not | total number of observations per year |
|---:|---:|---|---:|
| 7 | 48 | TRUE | 48 |

```
## further visualisation for determining whether the data is balanced

Fatalities$year <- as.factor(Fatalities$year)
balance_check <- Fatalities%>%
  select(state, year)%>%
  group_by(state,year)

state_count <- Fatalities%>%
  select(state)%>%
  group_by(state)

year_count <- Fatalities%>%
  select(year)%>%
  group_by(year)

nrow(summarise(year_count))
```

```
## [1] 7
```

```
nrow(summarise(state_count))
```

```
## [1] 48
```

```
table(year_count)
```

```
## year
## 1982 1983 1984 1985 1986 1987 1988
##   48   48   48   48   48   48   48
```

```
table(state_count)
```

```
## state
## al az ar ca co ct de fl ga id il in ia ks ky la me md ma mi mn ms mo mt ne nv
##  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
## nh nj nm ny nc nd oh ok or pa ri sc sd tn tx ut vt va wa wv wi wy
##  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
```

```
table(balance_check)
```

```
##      year
## state 1982 1983 1984 1985 1986 1987 1988
##    al    1    1    1    1    1    1    1
##    az    1    1    1    1    1    1    1
##    ar    1    1    1    1    1    1    1
##    ca    1    1    1    1    1    1    1
##    co    1    1    1    1    1    1    1
##    ct    1    1    1    1    1    1    1
##    de    1    1    1    1    1    1    1
##    fl    1    1    1    1    1    1    1
##    ga    1    1    1    1    1    1    1
##    id    1    1    1    1    1    1    1
##    il    1    1    1    1    1    1    1
##    in    1    1    1    1    1    1    1
##    ia    1    1    1    1    1    1    1
##    ks    1    1    1    1    1    1    1
##    ky    1    1    1    1    1    1    1
##    la    1    1    1    1    1    1    1
##    me    1    1    1    1    1    1    1
##    md    1    1    1    1    1    1    1
##    ma    1    1    1    1    1    1    1
##    mi    1    1    1    1    1    1    1
##    mn    1    1    1    1    1    1    1
##    ms    1    1    1    1    1    1    1
##    mo    1    1    1    1    1    1    1
##    mt    1    1    1    1    1    1    1
##    ne    1    1    1    1    1    1    1
##    nv    1    1    1    1    1    1    1
##    nh    1    1    1    1    1    1    1
##    nj    1    1    1    1    1    1    1
##    nm    1    1    1    1    1    1    1
##    ny    1    1    1    1    1    1    1
##    nc    1    1    1    1    1    1    1
##    nd    1    1    1    1    1    1    1
##    oh    1    1    1    1    1    1    1
##    ok    1    1    1    1    1    1    1
##    or    1    1    1    1    1    1    1
##    pa    1    1    1    1    1    1    1
##    ri    1    1    1    1    1    1    1
##    sc    1    1    1    1    1    1    1
##    sd    1    1    1    1    1    1    1
##    tn    1    1    1    1    1    1    1
##    tx    1    1    1    1    1    1    1
##    ut    1    1    1    1    1    1    1
##    vt    1    1    1    1    1    1    1
##    va    1    1    1    1    1    1    1
##    wa    1    1    1    1    1    1    1
##    wv    1    1    1    1    1    1    1
##    wi    1    1    1    1    1    1    1
##    wy    1    1    1    1    1    1    1
```

From our analysis, we can identify 7 years and 48 states. Further research suggests that the years are from 1982 to 1988, and the states are all contiguous federal states from the United States. As we can see from the data frame above, for all years, the total number of observations per year are 48. This suggests that the data is balanced since all variables are observed for all entities and over all time periods.

c. Now add a full set of state dummies into the basic model (denoted as model 2). Does the coefficient and significance level of BeerTaxit change? Please explain, using your own language, why there is a change (or lack of it).

```
## creating the linear model

model2 <- lm(fatal_rate ~ beertax + state, data = Fatalities)
summary(model2)
```

```
##
## Call:
## lm(formula = fatal_rate ~ beertax + state, data = Fatalities)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.870e-05 -8.284e-06 -1.270e-07  7.955e-06  8.978e-05
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.478e-04  3.134e-05  11.098  < 2e-16 ***
## beertax     -6.559e-05  1.879e-05  -3.491 0.000556 ***
## stateaz     -5.677e-05  2.667e-05  -2.129 0.034107 *
## statear     -6.549e-05  2.190e-05  -2.990 0.003028 **
## stateca     -1.509e-04  3.043e-05  -4.960 1.21e-06 ***
## stateco     -1.484e-04  2.873e-05  -5.165 4.50e-07 ***
## statect     -1.862e-04  2.805e-05  -6.638 1.58e-10 ***
## statede     -1.308e-04  2.939e-05  -4.448 1.24e-05 ***
## statefl     -2.681e-05  1.393e-05  -1.924 0.055284 .
## statega      5.246e-05  1.839e-05   2.852 0.004661 **
## stateid     -6.690e-05  2.580e-05  -2.593 0.009989 **
## stateil     -1.962e-04  2.915e-05  -6.730 9.23e-11 ***
## statein     -1.462e-04  2.725e-05  -5.363 1.69e-07 ***
## stateia     -1.544e-04  2.534e-05  -6.092 3.58e-09 ***
## stateks     -1.223e-04  2.454e-05  -4.984 1.08e-06 ***
## stateky     -1.218e-04  2.872e-05  -4.240 3.02e-05 ***
## statela     -8.471e-05  1.887e-05  -4.490 1.03e-05 ***
## stateme     -1.108e-04  1.911e-05  -5.797 1.78e-08 ***
## statemd     -1.706e-04  2.832e-05  -6.025 5.17e-09 ***
## statema     -2.110e-04  2.761e-05  -7.641 3.24e-13 ***
## statemi     -1.485e-04  2.360e-05  -6.290 1.18e-09 ***
## statemn     -1.897e-04  2.651e-05  -7.157 6.92e-12 ***
## statems     -2.908e-06  1.484e-05  -0.196 0.844839
## statemo     -1.296e-04  2.667e-05  -4.861 1.93e-06 ***
## statemt     -3.604e-05  2.640e-05  -1.365 0.173225
## statene     -1.522e-04  2.493e-05  -6.106 3.30e-09 ***
## statenv     -6.008e-05  2.859e-05  -2.101 0.036517 *
## statenh     -1.254e-04  2.097e-05  -5.983 6.53e-09 ***
## statenj     -2.106e-04  3.072e-05  -6.855 4.37e-11 ***
## statenm      4.264e-05  2.543e-05   1.677 0.094724 .
## stateny     -2.187e-04  2.989e-05  -7.316 2.57e-12 ***
## statenc     -2.905e-05  1.198e-05  -2.424 0.015979 *
## statend     -1.623e-04  2.538e-05  -6.396 6.45e-10 ***
## stateoh     -1.674e-04  2.538e-05  -6.597 2.02e-10 ***
## stateok     -5.451e-05  1.691e-05  -3.223 0.001415 **
## stateor     -1.168e-04  2.857e-05  -4.088 5.65e-05 ***
## statepa     -1.767e-04  2.761e-05  -6.402 6.26e-10 ***
## stateri     -2.265e-04  2.938e-05  -7.711 2.07e-13 ***
## statesc      5.572e-05  1.100e-05   5.065 7.30e-07 ***
## statesd     -1.004e-04  2.096e-05  -4.788 2.70e-06 ***
## statetn     -8.757e-05  2.680e-05  -3.267 0.001218 **
## statetx     -9.175e-05  2.456e-05  -3.736 0.000225 ***
## stateut     -1.164e-04  1.964e-05  -5.926 8.88e-09 ***
## statevt     -9.660e-05  2.111e-05  -4.576 7.08e-06 ***
## stateva     -1.290e-04  2.042e-05  -6.320 9.99e-10 ***
## statewa     -1.660e-04  2.835e-05  -5.854 1.31e-08 ***
## statewv     -8.967e-05  2.466e-05  -3.636 0.000328 ***
## statewi     -1.759e-04  2.939e-05  -5.985 6.44e-09 ***
## statewy     -2.285e-05  3.129e-05  -0.730 0.465811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.899e-05 on 287 degrees of freedom
## Multiple R-squared:  0.905,  Adjusted R-squared:  0.8891
## F-statistic: 56.97 on 48 and 287 DF,  p-value: < 2.2e-16
```

The coefficient of BeerTax change from 3.646e-05 in model1 to -6.559e-05 in model2. In both models, BeerTax are still statistically significant (p-value = 1.08e-08 in model1 and 0.000556 in model2 < 0.05), which suggests that its effects on fatality rate are still significantly different from 0.

As we can see, the coefficient of BeerTax changed from a positive value of 3.646e-05 to a negative value of -6.559e-05 in the new model. This change is due to us creating dummy variables and taking into account all time invariant variables across different states as it is a potential confounding variable. For example, different states have different cultures, density of cars and beer tax rates which might all lead to biased estimates for our dependent variable. As the difference in states is controlled, we can notice that over time, beertax leads to a decreased fatality rate from traffic incidence.

d. Now add a full set of year dummies into the basic model (denoted as model 3). Does the coefficient and significance level of BeerTaxit change? Please explain, using your own language, why there is a change (or lack of it).

```r
## creating the linear model

model3 <- lm(fatal_rate ~ beertax + year, data = Fatalities)
summary(model3)
```

```
##
## Call:
## lm(formula = fatal_rate ~ beertax + year, data = Fatalities)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.061e-04 -3.843e-05 -9.853e-06  2.678e-05  2.234e-04
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.895e-04  8.566e-06  22.121  < 2e-16 ***
## beertax      3.663e-05  6.260e-06   5.852 1.18e-08 ***
## year1983    -8.204e-06  1.117e-05  -0.735    0.463
## year1984    -7.173e-06  1.117e-05  -0.642    0.521
## year1985    -1.105e-05  1.117e-05  -0.990    0.323
## year1986    -1.612e-06  1.117e-05  -0.144    0.885
## year1987    -1.554e-06  1.117e-05  -0.139    0.889
## year1988    -1.027e-07  1.117e-05  -0.009    0.993
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.471e-05 on 328 degrees of freedom
## Multiple R-squared:  0.09865,    Adjusted R-squared:  0.07941
## F-statistic: 5.128 on 7 and 328 DF,  p-value: 1.499e-05
```

The coefficient of BeerTax change from 3.646e-05 in model1 to 3.663e-05 in model3. As the change in both models are minimal (1.7e-7), we can conclude that there is not much change in the coefficient of BeerTax in both models. In both models, BeerTax are still statistically significant (p-value = 1.08e-08 in model1 and 1.18e-08 in model2 < 0.05), which suggests that its effects on fatality rate are still different from 0.

As we can see, there is not much change in the coefficient of BeerTax. From this, together with the statistically insignificant coefficients of all the year dummy variables, we can see that the change in years does not affect the fatality rate in a significant way. This might be because all the years are back to back, and it is extremely hard for a city or a district to change their culture in a short period of time (for example in a few years). From this, we can also see that year is not a confounding variable, and including it into the model does not improve the model in any significant ways.

e. Now use the package plm in R to estimate a model with state fixed effects (denoted as model 4). Compare the coefficients in model 2 and model 4. Are the results identical? Why or why not?

```r
## creating the linear model

model4 <- plm(fatal_rate ~ beertax, data = Fatalities, effect = "individual", model = "within")
summary(model4)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = fatal_rate ~ beertax, data = Fatalities, effect = "individual",
##     model = "within")
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -5.8696e-05 -8.2838e-06 -1.2701e-07  7.9545e-06  8.9780e-05
##
## Coefficients:
##            Estimate  Std. Error t-value Pr(>|t|)
## beertax -6.5587e-05  1.8785e-05 -3.4915 0.000556 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.0785e-07
## Residual Sum of Squares: 1.0345e-07
## R-Squared:      0.040745
## Adj. R-Squared: -0.11969
## F-statistic: 12.1904 on 1 and 287 DF, p-value: 0.00055597
```

The coefficient of BeerTax remained the same at -6.5587e-05 for both model2 and model4. In both models, BeerTax are still statistically significant (p-value = 1.08e-08 in model2 and 0.000556 in model4 < 0.05), which suggests that its effects on fatality rate are still different from 0.

As we can see, the coefficient of BeerTax remained the same throughout the two models. This is due to both models performing the same entity fixed effect regression on the dataset. For model2, it utilises the “n-1 binary regressors” OLS regression, which is another notation of entity fixed effect regression that basically achieves the same outcome as entity fixed effect regression used above. Thus there are no doubt that the results are unchanged in both models.

f. Now use the package plm in R to estimate the basic model with year-fixed effects (denoted as model 5). Compare the coefficients in model 3 and model 5. Are the results identical? Why or why not?

```
## creating the linear model

model5 <- plm(fatal_rate ~ beertax, data = Fatalities, effect = "time", model = "within")
summary(model5)
```

```
## Oneway (time) effect Within Model
##
## Call:
## plm(formula = fatal_rate ~ beertax, data = Fatalities, effect = "time",
##     model = "within")
##
## Balanced Panel: n = 48, T = 7, N = 336
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -1.0607e-04 -3.8427e-05 -9.8534e-06  2.6781e-05  2.2345e-04
##
## Coefficients:
##           Estimate Std. Error t-value  Pr(>|t|)
## beertax 3.6634e-05 6.2600e-06   5.852 1.177e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1.0842e-06
## Residual Sum of Squares: 9.8171e-07
## R-Squared:      0.094538
## Adj. R-Squared: 0.075214
## F-statistic: 34.246 on 1 and 328 DF, p-value: 1.177e-08
```

The coefficient of BeerTax remained the same at 3.6634e-05 for both model3 and model5. In both models, BeerTax are still statistically significant (p-value = 1.177e-08 in model5 and 1.18e-08 in model3 < 0.05), which suggests that its effects on fatality rate are still different from 0.

As we can see, the coefficient of BeerTax remained the same. This is mainly due to both model3 and model5 performing the same time fixed effect regression on the dataset. For model3, it utilises the “t-1 binary regressors” OLS regression, which is another notation of time fixed effect regression that basically achieves the same outcome as the time fixed effect regression used above. Thus it is expected that the results are unchanged in both models. From this, we can see once again that the change in years does not affect the fatality rate in a significant way. This might be because all the years are back to back, and it is extremely hard for a city or a district to change the culture in a short period of time (for example in a few years).

g. A scholar argues that the state of the nation’s economy will affect the quality of road infrastructure and thus influence traffic fatalities. Supposing we have the data of GDP (Gross Domestic Product) for every year, do we need to add that variable into model 3 to control for the national economic condition? Why or why not?

From our team’s perspective, we believe that we shouldn’t include GDP into model 3. This is because model3 is a time fixed effect. For the national economic condition as GDP(a national index which is identical to all of the states in US), as it is a entity invariant variable, it doesn’t varies across entities. As it is already captured in our regressor in model 3, we do not need to control for it again as it is already controlled in the original regression model.

h. Follow the steps for a state fixed effects regression via demeaning and estimate a new model on the entity-demeaned variables. Are the results identical to model 4? Are they identical to model 2?

```
## creating the mean for entity-demeaned variables, including fatal_rate and beertax

fatalities_copy <- Fatalities

fatalities_copy <- fatalities_copy %>% group_by(state) %>% mutate(across(fatal_rate, ~ mean(.), .names = "{col}.m
ean"))
fatalities_copy <- fatalities_copy %>% group_by(state) %>% mutate(across(beertax, ~ mean(.), .names = "{col}.mean
"))
## fatalities_copy

fatalities_copy <- fatalities_copy %>%
  mutate(fatal_rate_demeaned = fatal_rate - fatal_rate.mean) %>%
  mutate(beertax_demeaned = beertax - beertax.mean)

model6 <- lm(fatal_rate_demeaned ~ beertax_demeaned, data = fatalities_copy)
summary(model6)
```

```
##
## Call:
## lm(formula = fatal_rate_demeaned ~ beertax_demeaned, data = fatalities_copy)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.870e-05 -8.284e-06 -1.270e-07  7.955e-06  8.978e-05
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -4.422e-21  9.601e-07   0.000 1.000000
## beertax_demeaned -6.559e-05  1.741e-05  -3.767 0.000196 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.76e-05 on 334 degrees of freedom
## Multiple R-squared:  0.04074,    Adjusted R-squared:  0.03787
## F-statistic: 14.19 on 1 and 334 DF,  p-value: 0.0001956
```

The coefficient of BeerTax is -6.559e-05 for model6, which is exactly the same as model2 and model4. In all three models, BeerTax are still statistically significant (p-value = 1.08e-08 in model2 and 0.000556 in model4 and 0.000196 in model6 < 0.05), which suggests that its effects on fatality rate are still different from 0.

As we can see, the coefficient of BeerTax remained the same throughout all three models. This is due to all three models performing the same entity fixed effect regression on the dataset. For model2, it utilises the “n-1 binary regressors” OLS regression, which is another notation of entity fixed effect regression that basically achieves the same outcome as entity fixed effect regression used in model4. Model6 utilises the “entity-demeaned” OLS regression, which essentially achieves the same outcome as model2 and model4 because it is just another way of representing and implementing the entity fixed effect regression. Thus there are no doubt that the results are unchanged in all three models.

i. Follow the steps for a year fixed effects regression via demeaning and estimate a model on the time-demeaned variables. Are the results identical to model 5? Are they identical to model 3?

```
## creating the mean for time-demeaned variables, including fatal_rate and beertax

fatalities_copy2 <- Fatalities

fatalities_copy2 <- fatalities_copy2 %>% group_by(year) %>% mutate(across(fatal_rate, ~ mean(.), .names = "{col}.mean"))
fatalities_copy2 <- fatalities_copy2 %>% group_by(year) %>% mutate(across(beertax, ~ mean(.), .names = "{col}.mean"))
fatalities_copy2
```

```
## # A tibble: 336 × 37
## # Groups:   year [7]
##    state year  spirits unemp income emppop beertax baptist mormon drinkage   dry
##    <fct> <fct>   <dbl> <dbl>  <dbl>  <dbl>   <dbl>   <dbl>  <dbl>    <dbl> <dbl>
##  1 al    1982     1.37 14.4  10544.   50.7   1.54    30.4   0.328     19    25.0
##  2 al    1983     1.36 13.7  10733.   52.1   1.79    30.3   0.343     19    23.0
##  3 al    1984     1.32 11.1  11109.   54.2   1.71    30.3   0.359     19    24.0
##  4 al    1985     1.28  8.90 11333.   55.3   1.65    30.3   0.376     19.7  23.6
##  5 al    1986     1.23  9.80 11662.   56.5   1.61    30.3   0.393     21    23.5
##  6 al    1987     1.18  7.80 11944    57.5   1.56    30.2   0.411     21    23.8
##  7 al    1988     1.17  7.20 12369.   56.8   1.50    30.2   0.430     21    23.8
##  8 az    1982     1.97  9.90 12309.   56.9   0.215    3.96  4.92      19     0
##  9 az    1983     1.90  9.10 12694.   57.6   0.206    3.89  4.83      19     0
## 10 az    1984     2.14  5    13266.   60.4   0.297    3.82  4.74      19     0
## # … with 326 more rows, and 26 more variables: youngdrivers <dbl>, miles <dbl>,
## #   breath <fct>, jail <fct>, service <fct>, fatal <int>, nfatal <int>,
## #   sfatal <int>, fatal1517 <int>, nfatal1517 <int>, fatal1820 <int>,
## #   nfatal1820 <int>, fatal2124 <int>, nfatal2124 <int>, afatal <dbl>,
## #   pop <dbl>, pop1517 <dbl>, pop1820 <dbl>, pop2124 <dbl>, milestot <dbl>,
## #   unempus <dbl>, emppopus <dbl>, gsp <dbl>, fatal_rate <dbl>,
## #   fatal_rate.mean <dbl>, beertax.mean <dbl>
```

```
fatalities_copy2 <- fatalities_copy2 %>%
  mutate(fatal_rate_demeaned = fatal_rate - fatal_rate.mean) %>%
  mutate(beertax_demeaned = beertax - beertax.mean)

model7 <- lm(fatal_rate_demeaned ~ beertax_demeaned, data = fatalities_copy2)
summary(model7)
```

```
##
## Call:
## lm(formula = fatal_rate_demeaned ~ beertax_demeaned, data = fatalities_copy2)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -1.061e-04 -3.843e-05 -9.853e-06  2.678e-05  2.234e-04
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.218e-21  2.958e-06   0.000        1
## beertax_demeaned 3.663e-05  6.204e-06   5.905 8.66e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.421e-05 on 334 degrees of freedom
## Multiple R-squared:  0.09454,    Adjusted R-squared:  0.09183
## F-statistic: 34.87 on 1 and 334 DF,  p-value: 8.663e-09
```

The coefficient of BeerTax remained at 3.663e-05 for model7, which is the same as model3 and model5 as stated above. In all three models, BeerTax are still statistically significant (p-value = 1.177e-08 in model5 and 1.18e-08 in model3 and 8.66e-09 in model7 < 0.05), which suggests that its effects on fatality rate are still different from 0.

As we can see, the coefficient of BeerTax remained the same. This is mainly due to all three models performing the same time fixed effect regression on the dataset. For model3, it utilises the “t-1 binary regressors” OLS regression, which is another notation of time fixed effect regression that basically achieves the same outcome as the time fixed effect regression used in model5. Futhermore, model7 utilises the “Time demeaned” OLS regression method, which is just another way of representing the time fixed effect regression mentioned above. Thus it is expected that the results are unchanged in all three models.