Traffic Fatalities

February 24, 2021

[1]:
```python
# Import required libraries.
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import ttest_ind
import statsmodels.api as sm
import statsmodels.formula.api as smformula
from linearmodels import PanelOLS
import scipy.stats as stats

%load_ext lab_black
```

[2]:
```python
# Read in the data:
data = pd.read_csv("./data/us_driving_fatalities.csv").iloc[:, 1:]
data.head()
```

[2]:
```
  state  year  spirits  unemp        income     emppop   beertax    baptist  \
0    al  1982     1.37   14.4  10544.152344  50.692039  1.539379  30.355700
1    al  1983     1.36   13.7  10732.797852  52.147030  1.788991  30.333599
2    al  1984     1.32   11.1  11108.791016  54.168087  1.714286  30.311501
3    al  1985     1.28    8.9  11332.626953  55.271137  1.652542  30.289499
4    al  1986     1.23    9.8  11661.506836  56.514496  1.609907  30.267401

    mormon  drinkage  …  nfatal2124      afatal         pop        pop1517  \
0  0.32829     19.00  …          32  309.437988  3942002.25  208999.593750
1  0.34341     19.00  …          35  341.834015  3960008.00  202000.078125
2  0.35924     19.00  …          34  304.872009  3988991.75  196999.968750
3  0.37579     19.67  …          45  276.742004  4021007.75  194999.734375
4  0.39311     21.00  …          29  360.716003  4049993.75  203999.890625

       pop1820        pop2124  milestot  unempus   emppopus       gsp
0  221553.43750  290000.06250   28516.0      9.7  57.799999 -0.022125
1  219125.46875  290000.15625   31032.0      9.6  57.900002  0.046558
2  216724.09375  288000.15625   32961.0      7.5  59.500004  0.062798
3  214349.03125  284000.31250   35091.0      7.2  60.100002  0.027490
4  212000.00000  263000.28125   36259.0      7.0  60.700001  0.032143
```

[5 rows x 34 columns]

[3]:
```
# Check dimensions:
Rows, Cols = data.shape
print("The data consists of {} rows and {} columns.".format(Rows, Cols))
```

```
The data consists of 336 rows and 34 columns.
```

0.0.1 Exercise 1.

How many states does this dataset contain? What's the time frame of this dataset? (From which year to which year). And what constitutes a single observation (i.e. what is the unit of analysis for each row of the data?)

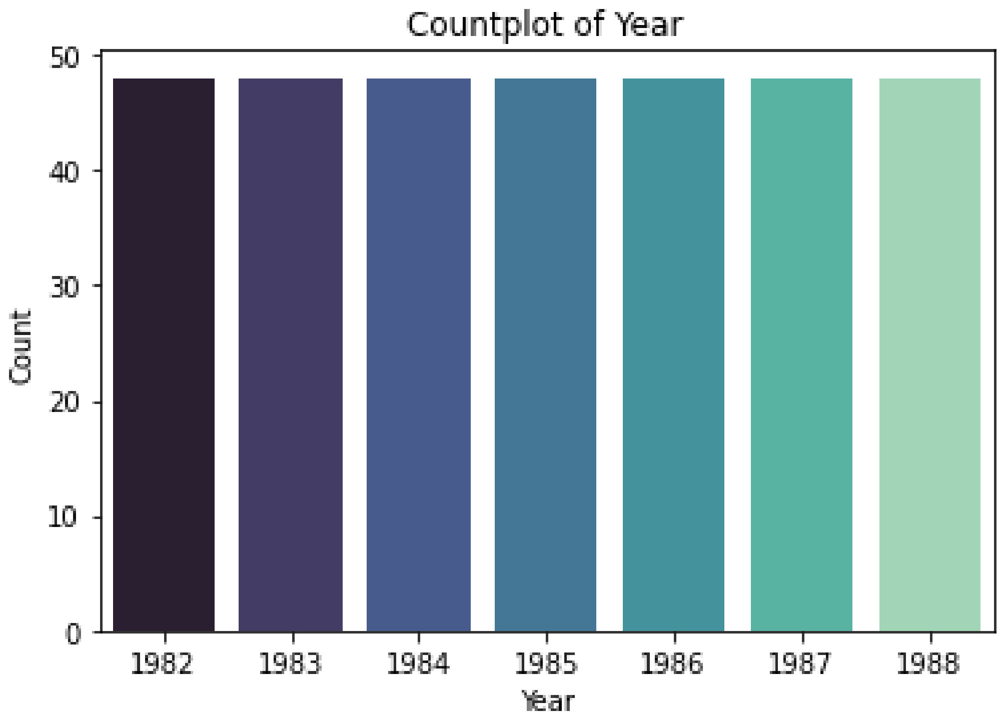
[4]:
```
States = list(set(data["state"]))
print("The data contains observations pertaining to {} states.".
 ↪format(len(States)))

minY = min(data["year"])
maxY = max(data["year"])
print(
    "The minimum year in the data is {}, and the maximum year is {}.".
 ↪format(minY, maxY)
)
```

```
The data contains observations pertaining to 48 states.
The minimum year in the data is 1982, and the maximum year is 1988.
```

> The unit of analysis for each row of the data, for now, is year-state-beertax (one row for one state-year). After we calculated the fatality rate and added other predictor variables, this will be expanded with the reponse variable and other covariates.

[5]:
```
Years = [str(i) for i in data["year"]]
sns.countplot(x=Years, palette="mako")
plt.xlabel("Year")
plt.ylabel("Count")
plt.title("Countplot of Year")
plt.show()
pass
```



0.0.2 Exercise 2.

We use the fatality rate per 10,000 as the dependent variable. Construct this variable. Name it as `fat_rate`. Hint: You can compute it using total fatalities (`fatal`) and population (`pop`). Note that because `pop` is often the name of a method in Python, you may have to navigate around some issues.

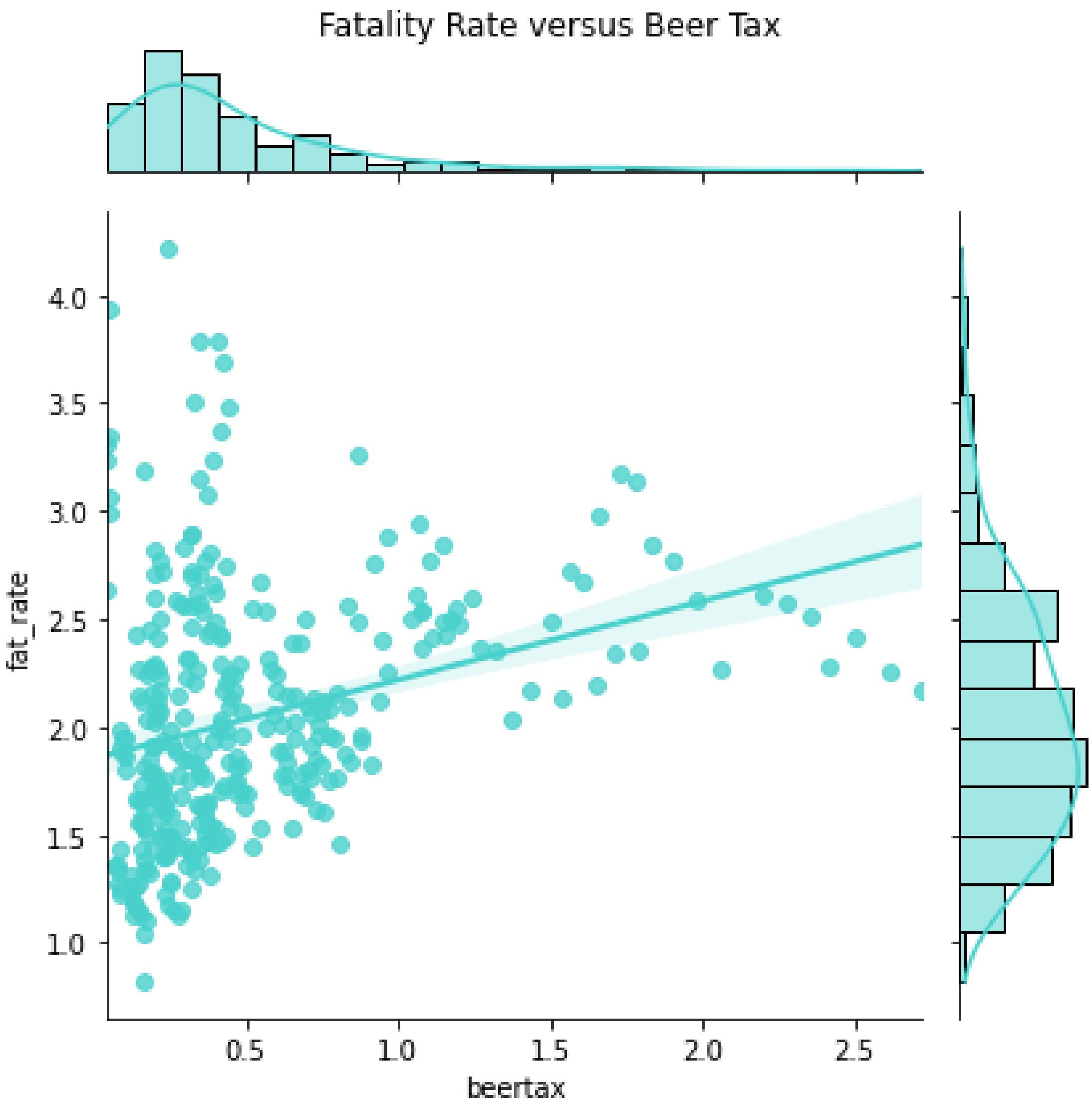
[6]:
```
Fatality_Rate = data["fatal"] / data["pop"] * 10000
data["fat_rate"] = Fatality_Rate
```

0.0.3 Exercise 3.

Draw a scatter plot using `beertax` as the x-axis, and `fat_rate` as the y-axis. Draw a fitted line showing the correlation between these two variables.

[7]:
```
p = sns.jointplot(
    x="beertax", y="fat_rate", data=data, kind="reg", color="mediumturquoise"
)
p.fig.suptitle("Fatality Rate versus Beer Tax")
p.fig.tight_layout()
p.fig.subplots_adjust(top=0.95)
# p.annotate(stats.pearsonr)
plt.xlabel("Beer Tax")
plt.ylabel("Fatality Rate, per 10,000")
```

```python
plt.show()
pass
```



0.0.4 Exercise 4.

Fit a simple OLS regression. This is what is called a "pooled" regression because we're "pooling" observations from different years into a single regression. What do your results imply about the relationship between Beer Taxes and fatalities?

$FatalityRate_i = \beta_0 + \beta_1 + BeerTax_i$

[8]:
```python
# Linear probability model; OLS. Use the statsmodels formula to code for␣
↪categorical variables.
```

```
FF4 = smformula.ols(formula="fat_rate ~ beertax", data=data).fit()
print(FF4.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               fat_rate   R-squared:                       0.093
Model:                            OLS   Adj. R-squared:                  0.091
Method:                 Least Squares   F-statistic:                     34.39
Date:                Wed, 24 Feb 2021   Prob (F-statistic):           1.08e-08
Time:                        16:45:20   Log-Likelihood:                -271.04
No. Observations:                 336   AIC:                             546.1
Df Residuals:                     334   BIC:                             553.7
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      1.8533      0.044     42.539      0.000       1.768       1.939
beertax        0.3646      0.062      5.865      0.000       0.242       0.487
==============================================================================
Omnibus:                       66.653   Durbin-Watson:                   0.465
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              112.734
Skew:                           1.134   Prob(JB):                     3.31e-25
Kurtosis:                       4.707   Cond. No.                         2.76
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

> With a small p value on the beertax variable ($p<0.05$), the regression summary shows that beertax is a statistically significant predictor variable for the fatality rate. However, the positive coefficient of beertax suggests that increasing beertax by one unit would lead to an increase in 0.36 unit of fatality rate on average, which is in contrast with our previous belief that increasing beer taxes could reduce alcohol consumption and could potenially reduce drunk driving deaths.

0.0.5 Exercise 5.

Now estimate your model again, this time adding state fixed effects (using the C() notation and your normal linear model machinery). What does this result imply about the relationship between beer taxes and fatalities?

[9]:
```
import warnings

warnings.filterwarnings("ignore")
```

```
# Linear probability model; OLS. Use the statsmodels formula to code for␣
↪categorical variables.
FF5 = smformula.ols(formula="fat_rate ~ beertax + C(state)", data=data).fit()
FF5.get_robustcov_results(cov_type="cluster", groups=data["state"]).summary()
```

[9]:

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:               fat_rate   R-squared:                       0.905
Model:                            OLS   Adj. R-squared:                  0.889
Method:                 Least Squares   F-statistic:                     9.944
Date:                Wed, 24 Feb 2021   Prob (F-statistic):            0.00281
Time:                        16:45:20   Log-Likelihood:                 107.97
No. Observations:                 336   AIC:                            -117.9
Df Residuals:                     287   BIC:                             69.09
Df Model:                          48
Covariance Type:              cluster
================================================================================
==
                     coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
--
Intercept          3.4776      0.511      6.802      0.000       2.449
4.506
C(state)[T.ar]    -0.6550      0.325     -2.013      0.050      -1.309
-0.001
C(state)[T.az]    -0.5677      0.413     -1.374      0.176      -1.399
0.264
C(state)[T.ca]    -1.5095      0.481     -3.139      0.003      -2.477
-0.542
C(state)[T.co]    -1.4843      0.451     -3.294      0.002      -2.391
-0.578
C(state)[T.ct]    -1.8623      0.438     -4.248      0.000      -2.744
-0.980
C(state)[T.de]    -1.3076      0.462     -2.828      0.007      -2.238
-0.377
C(state)[T.fl]    -0.2681      0.160     -1.676      0.100      -0.590
0.054
C(state)[T.ga]     0.5246      0.257      2.040      0.047       0.007
1.042
C(state)[T.ia]    -1.5439      0.389     -3.967      0.000      -2.327
-0.761
C(state)[T.id]    -0.6690      0.398     -1.683      0.099      -1.469
0.131
C(state)[T.il]    -1.9616      0.458     -4.283      0.000      -2.883
```

						-1.040
C(state)[T.in]	-1.4615	0.424	-3.447	0.001	-2.314	-0.609
C(state)[T.ks]	-1.2232	0.375	-3.266	0.002	-1.977	-0.470
C(state)[T.ky]	-1.2175	0.450	-2.704	0.010	-2.123	-0.312
C(state)[T.la]	-0.8471	0.267	-3.178	0.003	-1.383	-0.311
C(state)[T.ma]	-2.1097	0.430	-4.902	0.000	-2.976	-1.244
C(state)[T.md]	-1.7064	0.443	-3.851	0.000	-2.598	-0.815
C(state)[T.me]	-1.1079	0.271	-4.082	0.000	-1.654	-0.562
C(state)[T.mi]	-1.4845	0.357	-4.157	0.000	-2.203	-0.766
C(state)[T.mn]	-1.8972	0.410	-4.622	0.000	-2.723	-1.071
C(state)[T.mo]	-1.2963	0.413	-3.136	0.003	-2.128	-0.465
C(state)[T.ms]	-0.0291	0.182	-0.160	0.873	-0.394	0.336
C(state)[T.mt]	-0.3604	0.408	-0.882	0.382	-1.182	0.461
C(state)[T.nc]	-0.2905	0.107	-2.719	0.009	-0.505	-0.076
C(state)[T.nd]	-1.6234	0.390	-4.163	0.000	-2.408	-0.839
C(state)[T.ne]	-1.5222	0.382	-3.989	0.000	-2.290	-0.754
C(state)[T.nh]	-1.2545	0.308	-4.079	0.000	-1.873	-0.636
C(state)[T.nj]	-2.1057	0.486	-4.333	0.000	-3.083	-1.128
C(state)[T.nm]	0.4264	0.391	1.091	0.281	-0.360	1.213
C(state)[T.nv]	-0.6008	0.448	-1.341	0.186	-1.502	0.301
C(state)[T.ny]	-2.1867	0.471	-4.640	0.000	-3.135	-1.239
C(state)[T.oh]	-1.6744	0.390	-4.294	0.000	-2.459	-0.890
C(state)[T.ok]	-0.5451	0.227	-2.404	0.020	-1.001	-0.089
C(state)[T.or]	-1.1680	0.448	-2.609	0.012	-2.069	-0.267

```
C(state)[T.pa]    -1.7675      0.430     -4.107      0.000      -2.633
-0.902
C(state)[T.ri]    -2.2651      0.462     -4.902      0.000      -3.195
-1.336
C(state)[T.sc]     0.5572      0.071      7.835      0.000       0.414
0.700
C(state)[T.sd]    -1.0037      0.307     -3.265      0.002      -1.622
-0.385
C(state)[T.tn]    -0.8757      0.416     -2.106      0.041      -1.712
-0.039
C(state)[T.tx]    -0.9175      0.375     -2.448      0.018      -1.671
-0.164
C(state)[T.ut]    -1.1640      0.282     -4.130      0.000      -1.731
-0.597
C(state)[T.va]    -1.2902      0.297     -4.345      0.000      -1.887
-0.693
C(state)[T.vt]    -0.9660      0.310     -3.113      0.003      -1.590
-0.342
C(state)[T.wa]    -1.6595      0.444     -3.741      0.000      -2.552
-0.767
C(state)[T.wi]    -1.7593      0.462     -3.805      0.000      -2.689
-0.829
C(state)[T.wv]    -0.8968      0.377     -2.380      0.021      -1.655
-0.139
C(state)[T.wy]    -0.2285      0.496     -0.461      0.647      -1.226
0.769
beertax           -0.6559      0.315     -2.083      0.043      -1.289
-0.022
==============================================================================
Omnibus:                       53.045   Durbin-Watson:                   1.517
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              219.863
Skew:                           0.585   Prob(JB):                     1.81e-48
Kurtosis:                       6.786   Cond. No.                         187.
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
"""
```

By adding state fixed effects, it turns out that excpept for NM and SC, all other states have negative coefficients. This result shows that when consider state as a group level effect, beer taxes and car assident fatality rates seem to be negatively related. When looking at the p-values across states, a majority of them are statistically significant (p-value <0.05). Increasing beertax by one unit would lead to an decrease in 0.66 unit of fatality rate per 10,000 on average, holding all else equal. This result supports our belief that increasing beer taxes could reduce alcohol consumption and could potentially reduce drunk driving deaths.

0.0.6 Exercise 6.

Explain why your results in Exercises 4 (without fixed effects) and Exercise 5 (with state fixed effects) look so different. What does this imply about states with high beer taxes?

> In exercise 4 where fixed effects are not included, we only look at the overall impact of beer taxes on fatality rate that's averaged across all states and all other variables. However, in exercise 5 where fixed effects are included, we now consider state as a group level effect, and the impact of beer taxes on fatality rate are only measured within the specified state. Since the amount of beer taxes can vary greatly across states, the results from before and after we add the fixed effects can also vary considerably. Some states could have high fatality rates in the beginning, so if we do not control the effect of state, we may attribute the problem of high fatality to having high beer taxes. This could imply that the car accident fatality rates in certain states with larger negative beertax coefficients may have decreased to a greater extent due to the high beer taxes in those states.

0.0.7 Exercise 7.

Implement the entity-demeaned approach to estimate the fixed-effects model by hand.

[10]:
```
# group by state and generate mean values for fatality rate and beer tax for␣
↪each state
meanfat = pd.DataFrame(data.groupby("state")["fat_rate"].mean())
beertax = pd.DataFrame(data.groupby("state")["beertax"].mean())

# concatenate two dataframes
staten = pd.concat([meanfat, beertax], axis=1)

state1 = staten.reset_index()
state1.columns = ["state", "state_fat_rate", "state_beertax"]

state1.head()
```

[10]:

	state	state_fat_rate	state_beertax
0	al	2.412627	1.623793
1	ar	2.435336	0.590575
2	az	2.705900	0.311040
3	ca	1.904977	0.096336
4	co	1.866981	0.192672

[11]:
```
# merge the mean values of fatality rate and beer tax by state with our␣
↪original data

merged = data.merge(state1, left_on="state", right_on="state")

# calculate how much fatality rate and beertax for each unit of observation␣
↪deviate from the mean fatality rate and beertax of that state
merged["fatamean"] = merged["fat_rate"] - merged["state_fat_rate"]
```

```
merged["beermean"] = merged["beertax"] - merged["state_beertax"]

# fit the regression
FF7 = smformula.ols(formula="fatamean ~ beermean", data=merged, missing="drop").
 ↪fit()
print(FF7.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               fatamean   R-squared:                       0.041
Model:                            OLS   Adj. R-squared:                  0.038
Method:                 Least Squares   F-statistic:                     14.19
Date:                Wed, 24 Feb 2021   Prob (F-statistic):           0.000196
Time:                        16:45:20   Log-Likelihood:                 107.97
No. Observations:                 336   AIC:                            -211.9
Df Residuals:                     334   BIC:                            -204.3
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept  -1.735e-17      0.010  -1.81e-15      1.000      -0.019       0.019
beermean      -0.6559      0.174     -3.767      0.000      -0.998      -0.313
==============================================================================
Omnibus:                       53.045   Durbin-Watson:                   1.517
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              219.863
Skew:                           0.585   Prob(JB):                     1.81e-48
Kurtosis:                       6.786   Cond. No.                         18.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

0.0.8 Exercise 8.

Fit the model with state fixed-effect using PanelOLS / lfe. Compare it to your by-hand output. Interpret the result.

[12]:
```
# Move state groups into highest level of multi-index, with old index in second␣
 ↪level.
# PanelOLS will then see the first level as the `entity` identifier.
MIndex = data.set_index(["state", data.index])
MIndex.head()
```

[12]:
```
         year  spirits  unemp        income     emppop   beertax    baptist  \
state
al    0  1982     1.37   14.4  10544.152344  50.692039  1.539379  30.355700
```

```
    1  1983     1.36   13.7  10732.797852  52.147030  1.788991  30.333599
    2  1984     1.32   11.1  11108.791016  54.168087  1.714286  30.311501
    3  1985     1.28    8.9  11332.626953  55.271137  1.652542  30.289499
    4  1986     1.23    9.8  11661.506836  56.514496  1.609907  30.267401

          mormon  drinkage        dry  …      afatal         pop  \
state                                  …
al     0  0.32829     19.00  25.006300  …  309.437988  3942002.25
       1  0.34341     19.00  22.994200  …  341.834015  3960008.00
       2  0.35924     19.00  24.042601  …  304.872009  3988991.75
       3  0.37579     19.67  23.633900  …  276.742004  4021007.75
       4  0.39311     21.00  23.464701  …  360.716003  4049993.75

               pop1517       pop1820       pop2124  milestot  unempus  \
state
al     0  208999.593750  221553.43750  290000.06250   28516.0      9.7
       1  202000.078125  219125.46875  290000.15625   31032.0      9.6
       2  196999.968750  216724.09375  288000.15625   32961.0      7.5
       3  194999.734375  214349.03125  284000.31250   35091.0      7.2
       4  203999.890625  212000.00000  263000.28125   36259.0      7.0

          emppopus       gsp  fat_rate
state
al     0  57.799999 -0.022125   2.12836
       1  57.900002  0.046558   2.34848
       2  59.500004  0.062798   2.33643
       3  60.100002  0.027490   2.19348
       4  60.700001  0.032143   2.66914

[5 rows x 34 columns]
```

[13]:
```python
mod = PanelOLS.from_formula("fat_rate ~ beertax + EntityEffects", data=MIndex)
mod.fit(cov_type="clustered", cluster_entity=True)
```

[13]:
```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:               fat_rate   R-squared:                        0.0407
Estimator:                   PanelOLS   R-squared (Between):             -0.3805
No. Observations:                 336   R-squared (Within):               0.0407
Date:                Wed, Feb 24 2021   R-squared (Overall):             -0.3775
Time:                        16:45:20   Log-likelihood                    107.97
Cov. Estimator:             Clustered
                                        F-statistic:                      12.190
Entities:                          48   P-value                           0.0006
Avg Obs:                       7.0000   Distribution:                   F(1,287)
Min Obs:                       7.0000
Max Obs:                       7.0000   F-statistic (robust):             5.1576
```

```
                                        P-value                      0.0239
Time periods:                       336 Distribution:              F(1,287)
Avg Obs:                         1.0000
Min Obs:                         1.0000
Max Obs:                         1.0000

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
beertax       -0.6559     0.2888    -2.2710     0.0239     -1.2243     -0.0874
==============================================================================

F-test for Poolability: 52.179
P-value: 0.0000
Distribution: F(47,287)

Included effects: Entity
PanelEffectsResults, id: 0x7fa2dc4b1040
```

The PanelOLS returns a negative coefficient for beertax, -0.66, which is the same as the result from our by-hand output for beertax. Both are statistically significant. This means that both results show a negative correlation between beertax and fatality rate.

0.0.9 Exercise 9.

Now (using PanelOLS or lfe) estimate a fixed effects model using the following specification. Add fixed effects for both the state and the year, as well as the other covariates you think are important X_{it}).

Explain (a) the type of phenomenon we control for by adding year fixed effects, and (b) your choice of covariates. Cluster the standard error at the state level. Interpret the result.

$FatalityRate_{it} = \beta BeerTax_{it} + STate_i + Year_t + epsilon_{it}$

[14]:
```
# Fit a linear regression with both fixed effects - state and year - and two␣
↪other covariates - income and youngdrivers

FF8 = smformula.ols(
    formula="fat_rate ~ beertax + C(state) + C(year) + income + youngdrivers",␣
↪data=data
).fit()
FF8.get_robustcov_results(cov_type="cluster", groups=data["state"]).summary()
```

[14]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:               fat_rate   R-squared:                       0.924
```

```
Model:                            OLS   Adj. R-squared:                  0.909
Method:                 Least Squares   F-statistic:                     11.99
Date:                Wed, 24 Feb 2021   Prob (F-statistic):           1.51e-09
Time:                        16:45:21   Log-Likelihood:                 145.66
No. Observations:                 336   AIC:                            -177.3
Df Residuals:                     279   BIC:                             40.26
Df Model:                          56
Covariance Type:              cluster
================================================================================
===
                      coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
---
Intercept           1.7219      0.825      2.087      0.042       0.062
3.382
C(state)[T.ar]     -0.5140      0.402     -1.278      0.208      -1.323
0.295
C(state)[T.az]     -0.7739      0.524     -1.476      0.147      -1.829
0.281
C(state)[T.ca]     -2.1888      0.673     -3.251      0.002      -3.543
-0.835
C(state)[T.co]     -1.9743      0.596     -3.313      0.002      -3.173
-0.775
C(state)[T.ct]     -2.9336      0.702     -4.176      0.000      -4.347
-1.521
C(state)[T.de]     -1.7851      0.614     -2.906      0.006      -3.021
-0.549
C(state)[T.fl]     -0.6955      0.287     -2.421      0.019      -1.273
-0.118
C(state)[T.ga]      0.1957      0.332      0.590      0.558      -0.472
0.864
C(state)[T.ia]     -1.7455      0.495     -3.524      0.001      -2.742
-0.749
C(state)[T.id]     -0.6049      0.486     -1.246      0.219      -1.582
0.372
C(state)[T.il]     -2.4763      0.611     -4.054      0.000      -3.705
-1.247
C(state)[T.in]     -1.5887      0.551     -2.886      0.006      -2.696
-0.481
C(state)[T.ks]     -1.5977      0.495     -3.226      0.002      -2.594
-0.601
C(state)[T.ky]     -1.1226      0.554     -2.027      0.048      -2.237
-0.008
C(state)[T.la]     -0.8665      0.325     -2.663      0.011      -1.521
-0.212
C(state)[T.ma]     -2.8800      0.632     -4.557      0.000      -4.152
```

```
-1.608
C(state)[T.md]     -2.4070      0.623     -3.864      0.000      -3.660
-1.154
C(state)[T.me]     -1.2402      0.350     -3.542      0.001      -1.945
-0.536
C(state)[T.mi]     -1.8864      0.469     -4.023      0.000      -2.830
-0.943
C(state)[T.mn]     -2.3041      0.550     -4.190      0.000      -3.410
-1.198
C(state)[T.mo]     -1.5627      0.532     -2.936      0.005      -2.633
-0.492
C(state)[T.ms]      0.2375      0.234      1.015      0.315      -0.233
0.708
C(state)[T.mt]     -0.3578      0.502     -0.713      0.480      -1.368
0.652
C(state)[T.nc]     -0.4055      0.149     -2.725      0.009      -0.705
-0.106
C(state)[T.nd]     -1.7121      0.483     -3.544      0.001      -2.684
-0.740
C(state)[T.ne]     -1.7629      0.487     -3.617      0.001      -2.743
-0.782
C(state)[T.nh]     -1.8808      0.472     -3.987      0.000      -2.830
-0.932
C(state)[T.nj]     -3.0420      0.740     -4.112      0.000      -4.530
-1.554
C(state)[T.nm]      0.4698      0.478      0.984      0.330      -0.491
1.431
C(state)[T.nv]     -1.1034      0.608     -1.815      0.076      -2.326
0.119
C(state)[T.ny]     -2.8410      0.665     -4.272      0.000      -4.179
-1.503
C(state)[T.oh]     -1.9510      0.503     -3.875      0.000      -2.964
-0.938
C(state)[T.ok]     -0.6973      0.297     -2.351      0.023      -1.294
-0.101
C(state)[T.or]     -1.3418      0.570     -2.356      0.023      -2.488
-0.196
C(state)[T.pa]     -2.0918      0.565     -3.699      0.001      -3.229
-0.954
C(state)[T.ri]     -2.6267      0.602     -4.363      0.000      -3.838
-1.416
C(state)[T.sc]      0.5530      0.088      6.273      0.000       0.376
0.730
C(state)[T.sd]     -0.9975      0.377     -2.649      0.011      -1.755
-0.240
C(state)[T.tn]     -0.8607      0.515     -1.671      0.101      -1.897
0.175
```

```
C(state)[T.tx]     -1.1942      0.487     -2.453      0.018      -2.173
-0.215
C(state)[T.ut]     -1.0747      0.344     -3.121      0.003      -1.768
-0.382
C(state)[T.va]     -1.7965      0.431     -4.172      0.000      -2.663
-0.930
C(state)[T.vt]     -1.1464      0.404     -2.838      0.007      -1.959
-0.334
C(state)[T.wa]     -2.0679      0.592     -3.495      0.001      -3.258
-0.877
C(state)[T.wi]     -2.0300      0.588     -3.450      0.001      -3.214
-0.846
C(state)[T.wv]     -0.7079      0.465     -1.523      0.134      -1.643
0.227
C(state)[T.wy]     -0.4032      0.622     -0.648      0.520      -1.655
0.848
C(year)[T.1983]    -0.0915      0.041     -2.212      0.032      -0.175
-0.008
C(year)[T.1984]    -0.1487      0.068     -2.179      0.034      -0.286
-0.011
C(year)[T.1985]    -0.2333      0.083     -2.814      0.007      -0.400
-0.067
C(year)[T.1986]    -0.1910      0.112     -1.702      0.095      -0.417
0.035
C(year)[T.1987]    -0.2528      0.134     -1.886      0.065      -0.522
0.017
C(year)[T.1988]    -0.2965      0.159     -1.865      0.068      -0.616
0.023
beertax            -0.5856      0.385     -1.523      0.134      -1.359
0.188
income              0.0001   5.25e-05      2.794      0.008    4.11e-05
0.000
youngdrivers        0.6740      1.234      0.546      0.588      -1.809
3.157
==============================================================================
Omnibus:                       12.768   Durbin-Watson:                   1.648
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               27.243
Skew:                           0.075   Prob(JB):                     1.21e-06
Kurtosis:                       4.387   Cond. No.                     2.51e+06
==============================================================================

Notes:
[1] Standard Errors are robust to cluster correlation (cluster)
[2] The condition number is large, 2.51e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

(a) By adding an additional fixed effects variable - year, now we also consider year as a group level effect, which means the we now measure the impact of beer taxes on fatality in each specific year from 1983 to 1988. The summary result shows negative coefficients throughout all 6 years and the p values for each year is below 0.05, which is statistically significant. This indicates that within each single year in the study, beer taxes had a negative impact on average fatality rate, which supports our belief that increasing beer taxes could reduce alcohol consumption and could potentially reduce drunk driving deaths. **(b)** We choose to add income as an additional covariate because a heavier beer tax could potentially affect those in the low income group more than those in the high income group. We choose to add youngdrivers as another covariate because young drivers have higher change of driving recklessly because of lack of experience/lifestyle etc., which could potentially be related with the outcome - fatality rate. From our results, income turns out to be a statistically significant predictor ($p<0.05$) but with only a small positive coefficient (0.0001), and youngdrivers, although with a relatively larger coefficient (0.6740), turns out to be a statistically insignificant predictor ($p>0.05$). According to the Notes section, Standard Errors are robust to cluster correlation at the state level. While our point estimates have dropped from 0.36 (FF4) to -0.59 (FF8), we have increased the size of our standard errors. The SE on beertax, has gone from 0.06 without clustering to 0.39 with clustering.