Sehgal_Umang_lab03

April 18, 2018

```
In [3]: import numpy as np
        import pandas as pd
        import statsmodels.formula.api as smf
        import matplotlib.pyplot as plt
        %matplotlib inline
```

```
In [4]: yrbs = pd.read_table('C:/Users/Umang/Downloads/yrbs.tsv.bz2')
        yrbs.head()
```

Out[4]:

	year	age	sex	meth	tv	state
0	2003	13	M	0	1	XX
1	2003	13	M	1	1	XX
2	2003	13	M	1	1	XX
3	2003	13	M	0	1	XX
4	2003	13	M	0	0	XX

0.0.1 1.1 load the data

```
In [5]: yrbs = pd.read_table('yrbs.tsv.bz2')
        print(yrbs.head())
        print(yrbs.shape)
        print("Unique years: \n",yrbs.year.unique())
        print("\nUnique Age: \n",yrbs.age.unique())
        print(yrbs.sex.unique())
        print(yrbs.meth.unique())
        print(yrbs.tv.unique())
        print(yrbs.state.unique())
```

```
   year  age sex  meth  tv state
0  2003   13   M     0   1    XX
1  2003   13   M     1   1    XX
2  2003   13   M     1   1    XX
3  2003   13   M     0   1    XX
4  2003   13   M     0   0    XX
(58077, 6)
Unique years:
 [2003 2005 2007 2009]
```

```
Unique Age:
 [13 14 15 16 17]
['M' 'F']
[0 1]
[1 0]
['XX' 'MT']
```

```
In [6]: print("Dimension :", np.shape(yrbs))
```

```
Dimension : (58077, 6)
```

```
In [7]: yrbs['before'] = np.where(yrbs['year'] <= 2005, 'before', 'after')
```

```
In [8]: yrbs.head()
```

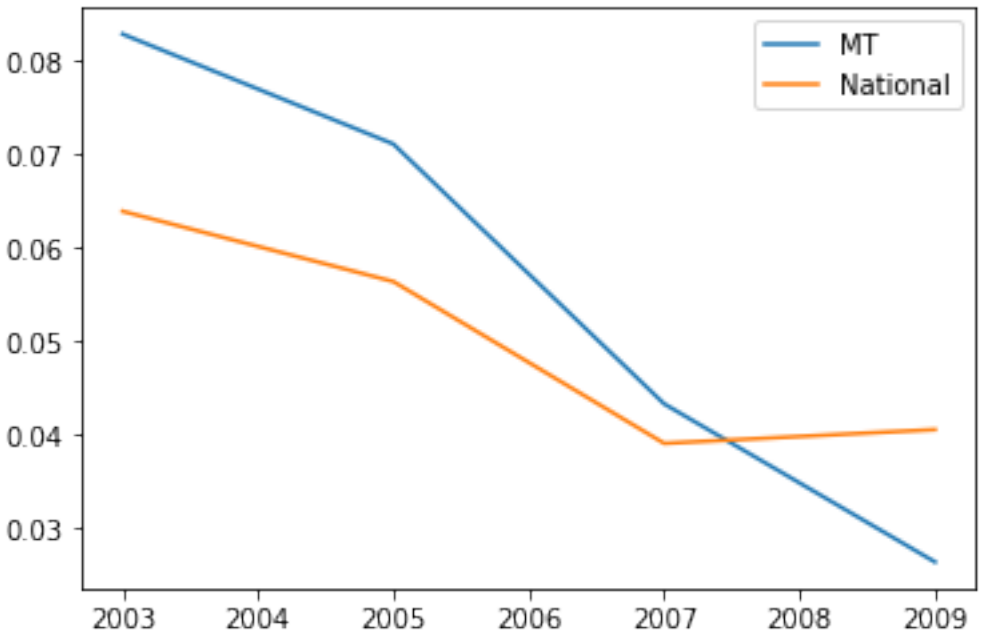
Out[8]:

	year	age	sex	meth	tv	state	before
0	2003	13	M	0	1	XX	before
1	2003	13	M	1	1	XX	before
2	2003	13	M	1	1	XX	before
3	2003	13	M	0	1	XX	before
4	2003	13	M	0	0	XX	before

0.0.2 1.2 graphical exploration

```
In [9]: ym = yrbs.groupby(['year', 'state']).meth.mean().reset_index()
        plt.plot(ym.year[ym.state=='MT'], ym.meth[ym.state=='MT'], label='MT')
        plt.plot(ym.year[ym.state=='XX'], ym.meth[ym.state=='XX'], label='National')
        plt.legend()
```

```
Out[9]: <matplotlib.legend.Legend at 0x20d0007f128>
```



0.0.3 1.3 Before-After

```
In [10]: print(yrbs[yrbs.state == 'MT'].groupby('before').meth.mean())
         m_bae = smf.ols(formula = 'meth ~ before', data = yrbs[yrbs.state == 'MT']).fit()

before
after     0.038115
before    0.076734
Name: meth, dtype: float64
```

There was a positive effect of camapign which reduced meth usage in Montana.

```
In [131]: m_bae.summary()

Out[131]: <class 'statsmodels.iolib.summary.Summary'>
          """
                                      OLS Regression Results
          ==============================================================================
          Dep. Variable:                   meth   R-squared:                       0.007
          Model:                            OLS   Adj. R-squared:                  0.007
          Method:                 Least Squares   F-statistic:                     67.66
          Date:                Wed, 18 Apr 2018   Prob (F-statistic):           2.19e-16
          Time:                        22:41:25   Log-Likelihood:                 417.76
          No. Observations:                9754   AIC:                            -831.5
          Df Residuals:                    9752   BIC:                            -817.1
          Df Model:                           1
          Covariance Type:            nonrobust
          ===================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
          -----------------------------------------------------------------------------------
          Intercept           0.0381      0.003     11.484      0.000       0.032       0.045
          before[T.before]    0.0386      0.005      8.225      0.000       0.029       0.048
          ==============================================================================
          Omnibus:                     7326.455   Durbin-Watson:                   1.957
          Prob(Omnibus):                  0.000   Jarque-Bera (JB):            84570.020
          Skew:                           3.765   Prob(JB):                         0.00
          Kurtosis:                      15.304   Cond. No.                         2.62
          ==============================================================================

          Warnings:
          [1] Standard Errors assume that the covariance matrix of the errors is correctly spe
          """
```

The Intercept of the above regression test confirms that after the proportion of meth users was 0.0381 as compared to (0.0381 + 0.0386) 0.0767 before the campaign.

```
In [11]: m_bae_full = smf.ols(formula = 'meth ~ before + age + sex + tv', data = yrbs[yrbs.sta

In [12]: m_bae_full.summary()

Out[12]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                  OLS Regression Results
         ==============================================================================
         Dep. Variable:                   meth   R-squared:                       0.009
         Model:                            OLS   Adj. R-squared:                  0.009
         Method:                 Least Squares   F-statistic:                     23.20
         Date:                Wed, 18 Apr 2018   Prob (F-statistic):           4.11e-19
         Time:                        23:41:07   Log-Likelihood:                 430.25
         No. Observations:                9754   AIC:                            -850.5
         Df Residuals:                    9749   BIC:                            -814.6
         Df Model:                           4
         Covariance Type:            nonrobust
         ==============================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
         ------------------------------------------------------------------------------
         Intercept          -0.1476      0.038     -3.911      0.000      -0.222      -0.074
         before[T.before]    0.0386      0.005      8.221      0.000       0.029       0.048
         sex[T.M]           -0.0031      0.005     -0.653      0.514      -0.012       0.006
         age                 0.0118      0.002      4.953      0.000       0.007       0.016
         tv                  0.0035      0.005      0.739      0.460      -0.006       0.013
         ==============================================================================
         Omnibus:                     7302.251   Durbin-Watson:                   1.963
         Prob(Omnibus):                  0.000   Jarque-Bera (JB):            83719.425
         Skew:                           3.750   Prob(JB):                         0.00
         Kurtosis:                      15.237   Cond. No.                         255.
         ==============================================================================

         Warnings:
         [1] Standard Errors assume that the covariance matrix of the errors is correctly spec
         """
```

We observe that when multiple variables or controls are added the effect changes for both Montana(drops by 0.1476) and nationally(drops by 0.109). That is there is a reverse impact.

1 Cross section estimator

```
In [13]: yrbs[yrbs.before == 'after'].groupby('state').meth.mean()

Out[13]: state
         MT    0.038115
         XX    0.039923
         Name: meth, dtype: float64

In [14]: m_cse = smf.ols(formula = 'meth ~ state', data = yrbs[yrbs.before=='after']).fit()
```

```
In [15]: m_cse.summary()

Out[15]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                   OLS Regression Results
         ==============================================================================
         Dep. Variable:                   meth   R-squared:                       0.000
         Model:                            OLS   Adj. R-squared:                 -0.000
         Method:                 Least Squares   F-statistic:                    0.3503
         Date:                Wed, 18 Apr 2018   Prob (F-statistic):             0.554
         Time:                        23:41:43   Log-Likelihood:                6404.0
         No. Observations:               29728   AIC:                         -1.280e+04
         Df Residuals:                   29726   BIC:                         -1.279e+04
         Df Model:                           1
         Covariance Type:            nonrobust
         ==============================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
         ------------------------------------------------------------------------------
         Intercept      0.0381      0.003     13.648      0.000       0.033       0.044
         state[T.XX]    0.0018      0.003      0.592      0.554      -0.004       0.008
         ==============================================================================
         Omnibus:                    27196.729   Durbin-Watson:                   1.976
         Prob(Omnibus):                  0.000   Jarque-Bera (JB):           619647.247
         Skew:                           4.720   Prob(JB):                         0.00
         Kurtosis:                      23.277   Cond. No.                         4.75
         ==============================================================================

         Warnings:
         [1] Standard Errors assume that the covariance matrix of the errors is correctly spec
         """
```

The regression test with no controls shows that for 0.0381 percent effect in Montana the nationwide effect of campaign for the reduction in meth use was 0.0399

```
In [16]: m_cse_full = smf.ols(formula = 'meth ~ state+  age + sex + tv    ', data = yrbs[yrbs.be

In [17]: m_cse_full.summary()

Out[17]: <class 'statsmodels.iolib.summary.Summary'>
         """
                                   OLS Regression Results
         ==============================================================================
         Dep. Variable:                   meth   R-squared:                       0.002
         Model:                            OLS   Adj. R-squared:                  0.001
         Method:                 Least Squares   F-statistic:                    11.60
         Date:                Wed, 18 Apr 2018   Prob (F-statistic):           2.07e-09
         Time:                        23:42:00   Log-Likelihood:                6427.0
         No. Observations:               29728   AIC:                         -1.284e+04
         Df Residuals:                   29723   BIC:                         -1.280e+04
```

```
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      -0.0649      0.018     -3.615      0.000      -0.100      -0.030
state[T.XX]     0.0015      0.003      0.494      0.621      -0.005       0.008
sex[T.M]        0.0081      0.002      3.563      0.000       0.004       0.013
age             0.0064      0.001      5.648      0.000       0.004       0.009
tv             -0.0016      0.002     -0.683      0.494      -0.006       0.003
==============================================================================
Omnibus:                    27150.346   Durbin-Watson:                   1.980
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           615983.002
Skew:                           4.709   Prob(JB):                         0.00
Kurtosis:                      23.214   Cond. No.                         253.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly spec
"""
```

On adding controls the values of intercepts changed for Montana the campaign effect reversed and meth use increased by a proportion of 0.0634.

Diiferences-in-Differences Estimator

```
In [19]: yrbs[yrbs.before == 'after'].groupby('state').meth.mean()

Out[19]: state
         MT    0.038115
         XX    0.039923
         Name: meth, dtype: float64

In [20]: yrbs[yrbs.before == 'before'].groupby('state').meth.mean()

Out[20]: state
         MT    0.076734
         XX    0.060319
         Name: meth, dtype: float64

In [21]: yrbs.groupby(['state', 'before']).meth.mean()

Out[21]: state  before
         MT     after     0.038115
                before    0.076734
         XX     after     0.039923
                before    0.060319
         Name: meth, dtype: float64

In [22]: print("Average use in Montana before and after the campaign : " , (abs(0.038115-0.076
```

```
Average use in Montana before and after the campaign :  0.038618999999999994
```

```
In [111]: print("\n Average use nationwide before and after the campaign : " , (abs(0.039923-0
```

```
 Average use nationwide before and after the campaign :  0.020395999999999997
```

```
In [23]: print("\n Trend difference : ", (abs(0.038618999999999994-0.020395999999999997)))
```

```
 Trend difference :  0.018222999999999996
```

```
In [25]: m_did = smf.ols(formula = 'meth ~ state*before  ', data = yrbs).fit()
```

```
In [26]: m_did.summary()
```

```
Out[26]: <class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                   meth   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.003
Method:                 Least Squares   F-statistic:                     62.92
Date:                Wed, 18 Apr 2018   Prob (F-statistic):           1.31e-40
Time:                        23:44:09   Log-Likelihood:                 5566.9
No. Observations:               58077   AIC:                         -1.113e+04
Df Residuals:                   58073   BIC:                         -1.109e+04
Df Model:                           3
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025
------------------------------------------------------------------------------------------------
Intercept                        0.0381      0.003     12.110      0.000       0.032
state[T.XX]                      0.0018      0.003      0.525      0.599      -0.005
before[T.before]                 0.0386      0.004      8.674      0.000       0.030
state[T.XX]:before[T.before]    -0.0182      0.005     -3.733      0.000      -0.028
==============================================================================
Omnibus:                    46583.747   Durbin-Watson:                   1.971
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           670089.493
Skew:                           4.057   Prob(JB):                         0.00
Kurtosis:                      17.528   Cond. No.                         12.2
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly speci
"""
```

Dif-in-dif regression tells the result of trend difference of simple test and confirms that the campaign had a reverse effect on meth use.

```
In [27]: m_did_full= smf.ols(formula = 'meth ~ state*before + age + sex + tv ', data = yrbs).fi

In [120]: m_did_full.summary()

Out[120]: <class 'statsmodels.iolib.summary.Summary'>
          """
                                      OLS Regression Results
          ==============================================================================
          Dep. Variable:                   meth   R-squared:                       0.005
          Model:                            OLS   Adj. R-squared:                  0.004
          Method:                 Least Squares   F-statistic:                     44.37
          Date:                Wed, 18 Apr 2018   Prob (F-statistic):           1.87e-54
          Time:                        20:55:51   Log-Likelihood:                 5605.5
          No. Observations:               58077   AIC:                         -1.120e+04
          Df Residuals:                   58070   BIC:                         -1.113e+04
          Df Model:                           6
          Covariance Type:            nonrobust
          ================================================================================
                                           coef    std err          t      P>|t|      [0.025
          --------------------------------------------------------------------------------
          Intercept                     -0.0727      0.015     -4.925      0.000      -0.102
          state[T.XX]                    0.0018      0.003      0.521      0.602      -0.005
          before[T.before]               0.0389      0.004      8.740      0.000       0.030
          sex[T.M]                       0.0064      0.002      3.486      0.000       0.003
          state[T.XX]:before[T.before]  -0.0185      0.005     -3.785      0.000      -0.028
          age                            0.0070      0.001      7.630      0.000       0.005
          tv                            -0.0041      0.002     -2.154      0.031      -0.008
          ==============================================================================
          Omnibus:                    46508.899   Durbin-Watson:                   1.973
          Prob(Omnibus):                  0.000   Jarque-Bera (JB):           666670.662
          Skew:                           4.049   Prob(JB):                         0.00
          Kurtosis:                      17.489   Cond. No.                         259.
          ==============================================================================

          Warnings:
          [1] Standard Errors assume that the covariance matrix of the errors is correctly spe
          """
```

Based on the full regression with all the controls we say that the campaign had efficacy in Montana as compared to the rest of the nation.