propensity_score_exercise_27FEB20

March 1, 2020

1 IDS 690: Unifying Data Science: Propensity Score Problem Set

[37]:
```
# Loading libraries
import pandas as pd
import numpy as np
import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols
from linearmodels import PanelOLS
import matplotlib.pyplot as plt
from scipy.stats import ttest_ind
```

[38]:
```
# Loading data
cps = pd.read_stata('https://github.com/nickeubank/MIDS_Data/blob/master/
↪Current_Population_Survey/morg18.dta?raw=true')

# Limit to people currently employed and working full time.
cps = cps[cps.lfsr94 == 'Employed-At Work']
cps = cps[cps.uhourse >= 35]

# And we can adjust earnings per hour (in cents) into dollars,
cps['earnhre_dollars'] = cps['earnhre'] / 100
cps['annual_earnings'] = cps['earnhre_dollars'] * cps['uhourse'] * 52

# And create gender and college educ variable
cps['female'] = (cps.sex == 2).astype('int')
cps['has_college_educ'] = (cps.grade92 > 43).astype('int')

cps.describe()
```

[38]:
```
              county       smsastat            age            sex  \
count  133814.000000  132638.000000  133814.000000  133814.000000
mean       25.735020       1.173932      43.335458       1.440320
std        61.578816       0.379052      13.335412       0.496427
min         0.000000       1.000000      16.000000       1.000000
25%         0.000000       1.000000      32.000000       1.000000
50%         0.000000       1.000000      43.000000       1.000000
```

```
75%          29.000000        1.000000       54.000000        2.000000
max         810.000000        2.000000       85.000000        2.000000

             grade92            race          ethnic         marital  \
count  133814.000000   133814.000000    18480.000000   133814.000000
mean       41.059680        1.434274        2.581872        3.253359
std         2.512128        1.270713        2.417939        2.676927
min        31.000000        1.000000        1.000000        1.000000
25%        39.000000        1.000000        1.000000        1.000000
50%        41.000000        1.000000        1.000000        1.000000
75%        43.000000        1.000000        4.000000        7.000000
max        46.000000       26.000000        8.000000        7.000000

             uhourse         earnhre        earnwke        chldpres  \
count  133814.000000    65755.000000  122603.000000   133814.000000
mean       42.596515     1940.998783    1105.295075        1.728451
std         7.002970     1008.707762     678.580654        3.053487
min        35.000000       17.000000       0.000000        0.000000
25%        40.000000     1300.000000     600.000000        0.000000
50%        40.000000     1675.000000     900.000000        0.000000
75%        40.000000     2300.000000    1403.840000        3.000000
max        99.000000     9999.000000    2884.610000       15.000000

            ownchild             ged      gedhigr         yrcoll  grprof  gr6cor  \
count  133814.000000    35118.000000  3107.000000   36240.000000     0.0     0.0
mean        0.630084        1.088473     6.640489       2.853256     NaN     NaN
std         1.028907        0.283986     1.321649       0.963869     NaN     NaN
min         0.000000        1.000000     1.000000       1.000000     NaN     NaN
25%         0.000000        1.000000     6.000000       2.000000     NaN     NaN
50%         0.000000        1.000000     7.000000       3.000000     NaN     NaN
75%         1.000000        1.000000     8.000000       3.000000     NaN     NaN
max        11.000000        2.000000     8.000000       5.000000     NaN     NaN

       ms123        occ2012  earnhre_dollars  annual_earnings         female  \
count    0.0  133814.000000     65755.000000     65755.000000  133814.000000
mean     NaN    3989.409128        19.409988     41757.890924       0.440320
std      NaN    2708.186730        10.087078     23164.092147       0.496427
min      NaN      10.000000         0.170000       397.800000       0.000000
25%      NaN    1550.000000        13.000000     27040.000000       0.000000
50%      NaN    4050.000000        16.750000     35360.000000       0.000000
75%      NaN    5700.000000        23.000000     49920.000000       1.000000
max      NaN    9750.000000        99.990000    361920.000000       1.000000

       has_college_educ
count     133814.000000
mean           0.148295
std            0.355394
```

```
min              0.000000
25%              0.000000
50%              0.000000
75%              0.000000
max              1.000000
```

1.1 Exercise 1

How many observations have a college degree, how many does not have a college degree.

[39]:
```
cps['has_college_educ'].value_counts()[0]
print(f'The sample includes {cps["has_college_educ"].value_counts()[0]}␣
↪observations that have not college education and {cps["has_college_educ"].
↪value_counts()[1]} observations that have college education.')
```

```
The sample includes 113970 observations that have not college education and
19844 observations that have college education.
```

1.2 Exercise 2

Show the raw difference of earnhre_dollars between the group with college degree and that without the college.

[40]:
```
# Calculating w/degree
cps_w_degree = cps.loc[(cps['has_college_educ'] == 1)]
cps_w_degree_mean = cps_w_degree['earnhre_dollars'].mean()

# W/O degree
cps_w_o_degree = cps.loc[(cps['has_college_educ'] == 0)]
cps_w_o_degree_mean = cps_w_o_degree['earnhre_dollars'].mean()

cps_diff = cps_w_degree_mean - cps_w_o_degree_mean

# Printing the difference
print(f' The mean difference between people with and without college degrees is␣
↪: {cps_diff:.2f}')
```

```
 The mean difference between people with and without college degrees is : 10.97
```

1.3 Exercise 3

Select the covariates that may be correlated with the treatment and dependent variables, use these covariates fit a logistic model to obtain propensity score.

We subset the covariates to those that we consider are not influenced by the effect of treatment. This is the case of county, age, sex, race and ethnic.

[41]:
```
logit = smf.logit('has_college_educ ~ county + age + sex + race + ethnic', cps).
↪fit()
```

```
logit.summary()
```

```
Optimization terminated successfully.
         Current function value: 0.225025
         Iterations 7
```

[41]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                           Logit Regression Results
==============================================================================
Dep. Variable:       has_college_educ   No. Observations:                18480
Model:                          Logit   Df Residuals:                    18474
Method:                           MLE   Df Model:                            5
Date:                Sat, 29 Feb 2020   Pseudo R-squ.:                 0.03870
Time:                        20:00:01   Log-Likelihood:                -4158.5
converged:                       True   LL-Null:                       -4325.9
Covariance Type:            nonrobust   LLR p-value:                 3.198e-70
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -4.8214      0.151    -31.953      0.000      -5.117      -4.526
county        -0.0003      0.001     -0.524      0.600      -0.001       0.001
age            0.0196      0.002      8.152      0.000       0.015       0.024
sex            0.5668      0.062      9.204      0.000       0.446       0.688
race           0.0329      0.022      1.523      0.128      -0.009       0.075
ethnic         0.1444      0.011     13.032      0.000       0.123       0.166
==============================================================================
"""
```

Calculate the propensity scores with the matcher package.

[42]:
```
from pymatch.Matcher import Matcher

pd.options.display.max_columns = 25
cps.groupby("has_college_educ").mean()

cps.columns
```

[42]:
```
Index(['county', 'smsastat', 'age', 'sex', 'grade92', 'race', 'ethnic',
       'marital', 'uhourse', 'earnhre', 'earnwke', 'chldpres', 'ownchild',
       'ged', 'gedhigr', 'yrcoll', 'grprof', 'gr6cor', 'ms123', 'occ2012',
       'lfsr94', 'class94', 'unioncov', 'ind02', 'stfips', 'earnhre_dollars',
       'annual_earnings', 'female', 'has_college_educ'],
      dtype='object')
```

[43]:
```
# Removed ged, gedhigr, yrcoll, grprof, gr6, ms123, Do I need earnhre and␣
↪earnhre_dollars or lfsr94? 'earnhre_dollars','annual_earnings',
```

```
#  'chldpres', 'class94', 'unioncov', 'ind02',  'female','smastat'
data = cps[['age', 'sex', 'race', 'ethnic','marital', 'uhourse',␣
 ↪'has_college_educ', "earnhre_dollars"]]
```

[44]:
```
# Creating the treatment variable
treatment = data[data.has_college_educ == 1]
control = data[data.has_college_educ == 0]
```

[45]:
```
# This shows a data inbalance.
m = Matcher(control, treatment, yvar="has_college_educ", ␣
 ↪exclude=['earnhre_dollars','marital', 'uhourse'])
```

```
Formula:
has_college_educ ~ age+sex+race+ethnic
n majority: 11465
n minority: 257
```

[46]:
```
np.random.seed(123)
m.fit_scores(balance=True, nmodels=250)
```

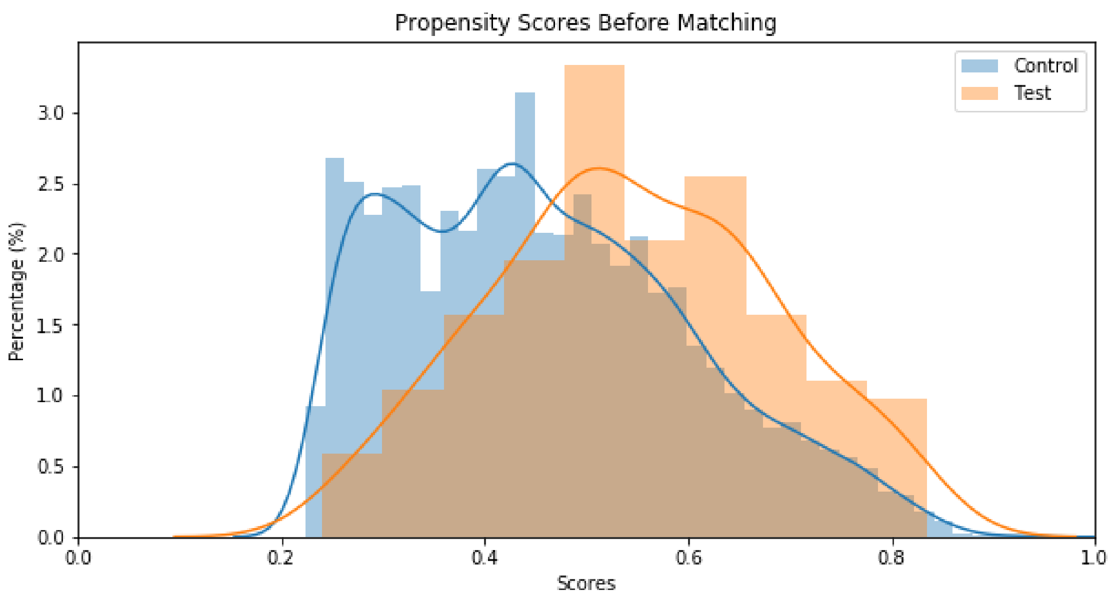
```
Fitting Models on Balanced Samples: 250\250
Average Accuracy: 62.45%
```

1.4 Exercise 4:

Finding common support.

[47]:
```
# Predicting Scores
m.predict_scores()
```

[48]:
```
# Plot Scores
m.plot_scores()
```



The plot present a significant portion of overlap between the treatment and control group, for the propensity scores between 0.2 and 0.8

1.5 Exercise 5

Obtain a matched sample using k:1 nearest neighbor method. Show the top ten rows of the matched data

[59]:
```
m.match(method="min", nmatches=1)
# Decent matching with propensity score
m.record_frequency()
```

[59]:
```
   freq  n_records
0     1        427
1     2         24
2     3          9
3     4          3
```

[50]:
```
# Generating the Weight Vector
m.assign_weight_vector()
```

[51]:
```
# Looking good
matched = m.matched_data.sort_values("match_id")
matched.head(10)
```

[51]:
```
     record_id    weight  age  sex  race  ethnic  marital  uhourse  \
35       18185  1.000000   31    2     3     1.0        3     40.0
257     113988  1.000000   31    2     3     1.0        3     40.0
```

```
46      21167  1.000000  45    1    1   7.0      1   40.0
258    114079  1.000000  45    1    1   7.0      1   40.0
51      21870  1.000000  35    1    1   1.0      7   40.0
259    114114  1.000000  35    1    1   1.0      1   40.0
260    114147  1.000000  26    1    1   1.0      7   40.0
123     58220  0.333333  26    1    1   1.0      1   40.0
133     60146  0.500000  43    2    1   1.0      1   40.0
261    114159  1.000000  43    2    1   1.0      1   40.0

     has_college_educ  earnhre_dollars    scores  match_id
35                  0            14.50  0.512419         0
257                 1            67.00  0.512419         0
46                  0            28.00  0.591440         1
258                 1            65.00  0.591440         1
51                  0            13.00  0.305958         2
259                 1            33.00  0.305958         2
260                 1            20.00  0.265071         3
123                 0            11.50  0.265071         3
133                 0            11.00  0.526724         4
261                 1            32.56  0.526724         4
```

[52]:
```
matched.groupby("has_college_educ").mean()
```

[52]:
```
                     record_id    weight        age       sex      race  \
has_college_educ
0                 57969.361868  0.801556  42.241245  1.579767  1.163424
1                124290.354086  1.000000  41.992218  1.587549  1.424125

                   ethnic   marital    uhourse  earnhre_dollars    scores  \
has_college_educ
0                3.540856  3.490272  40.190661        17.483502  0.545487
1                3.381323  3.548638  40.949416        26.867782  0.545492

                 match_id
has_college_educ
0                   128.0
1                   128.0
```

1.6 Exercise 6:

Conduct a t-test between the treatment and control group using the matched data. Interpret the result. Are covariates balanced?

[74]:
```
# Matching the two DFs
# Conduct a t-test between the treatment and control group using the matched␣
↪data. Interpret the result. Are covariates balanced?
matched_w_college = matched.loc[(matched['has_college_educ'] == 1)]
```

```
matched_wo_college = matched.loc[(matched['has_college_educ'] == 0)]
```

[75]:
```
#Check the statistical significance of mean difference across selected␣
↪covariates in the matcher.
covariates = ['age','sex','race','ethnic','scores']

for i in covariates:
    print(f'P_value from t-test on {i}:␣
↪{ttest_ind(matched_w_college[i],matched_wo_college[i])[1]:.2f}')
```

```
P_value from t-test on age: 0.80
P_value from t-test on sex: 0.86
P_value from t-test on race: 0.02
P_value from t-test on ethnic: 0.50
P_value from t-test on scores: 1.00
```

The p_value for race is below 0.05, which may indicate insuficient balance for this covariate.

1.7 Exercise 7:

Fit four separate regression models to estimate the effect of college education on earning per hour. - an OLS model, including only the treatment variable - an OLS model, including the treatment variable and covariates - a weighted least squared model, including only the treatment variable, using the weight obtained by propensity score matching - a weighted least squared model, including the treatment variable and covariates, using the weight obtained by propensity score matching

Compare the above four models, interpret the results.

Fitting a OLS model that only controls for the effect of treatment (has college), we obtain that the mean difference is $10.96

[76]:
```
# Fitting a Model w/only response and treatment which raises hourly income by␣
↪$10.96 per hour.
smf.ols('earnhre_dollars ~ has_college_educ', data = cps).fit().summary()
```

[76]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:        earnhre_dollars   R-squared:                       0.052
Model:                            OLS   Adj. R-squared:                  0.052
Method:                 Least Squares   F-statistic:                     3577.
Date:                Sun, 01 Mar 2020   Prob (F-statistic):               0.00
Time:                        22:08:18   Log-Likelihood:            -2.4354e+05
No. Observations:               65755   AIC:                         4.871e+05
Df Residuals:                   65753   BIC:                         4.871e+05
Df Model:                           1
Covariance Type:            nonrobust
```

```
==============================================================================
====
                       coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----
Intercept           18.9084      0.039    482.151      0.000      18.832
18.985
has_college_educ    10.9654      0.183     59.804      0.000      10.606
11.325
==============================================================================
Omnibus:                28321.811   Durbin-Watson:                   1.878
Prob(Omnibus):              0.000   Jarque-Bera (JB):           167903.335
Skew:                       2.001   Prob(JB):                         0.00
Kurtosis:                   9.728   Cond. No.                         4.80
==============================================================================

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

Fitting a OLS model adding controls for demographic information of treatment (has college), the mean difference reduces to $9.91

[77]:
```
# Fitting an OLS model, including the treatment variable and covariates␣
↪(treatment raises pay by $9.91)
smf.ols('earnhre_dollars ~ has_college_educ + age + sex + C(race) + C(ethnic)',␣
↪data = cps).fit().summary()
```

[77]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:        earnhre_dollars   R-squared:                       0.083
Model:                            OLS   Adj. R-squared:                  0.080
Method:                 Least Squares   F-statistic:                     38.96
Date:                Sun, 01 Mar 2020   Prob (F-statistic):         4.20e-195
Time:                        22:08:20   Log-Likelihood:                -40529.
No. Observations:               11722   AIC:                         8.111e+04
Df Residuals:                   11694   BIC:                         8.132e+04
Df Model:                          27
Covariance Type:            nonrobust
==============================================================================
====
                       coef    std err          t      P>|t|      [0.025
0.975]
```

```
--------------------------------------------------------------------------------
----
Intercept          16.7399      0.310     54.057      0.000      16.133
17.347
C(race)[T.2]       -0.2745      0.406     -0.677      0.499      -1.070
0.521
C(race)[T.3]        0.8258      0.538      1.536      0.125      -0.228
1.880
C(race)[T.4]        1.2683      0.837      1.515      0.130      -0.372
2.909
C(race)[T.5]        0.4750      1.082      0.439      0.661      -1.647
2.597
C(race)[T.6]        1.0664      0.930      1.147      0.251      -0.756
2.889
C(race)[T.7]       -0.2413      0.674     -0.358      0.720      -1.562
1.080
C(race)[T.8]       -1.2643      1.988     -0.636      0.525      -5.162
2.633
C(race)[T.9]        6.2158      3.847      1.616      0.106      -1.324
13.756
C(race)[T.10]       0.9967      2.135      0.467      0.641      -3.189
5.182
C(race)[T.11]      -3.1340      7.690     -0.408      0.684     -18.208
11.940
C(race)[T.12]    -4.89e-13   4.67e-13     -1.048      0.295    -1.4e-12
4.26e-13
C(race)[T.13]       6.2430      7.691      0.812      0.417      -8.832
21.318
C(race)[T.14]       3.2723      7.695      0.425      0.671     -11.810
18.355
C(race)[T.15]      -1.2323      3.444     -0.358      0.721      -7.983
5.519
C(race)[T.16]       0.7416      2.570      0.289      0.773      -4.297
5.780
C(race)[T.17]       3.8199      7.701      0.496      0.620     -11.275
18.915
C(race)[T.18]   -1.697e-15   3.38e-15     -0.502      0.616   -8.33e-15
4.93e-15
C(race)[T.19]    3.072e-15   4.78e-15      0.643      0.520    -6.3e-15
1.24e-14
C(race)[T.20]   -8.696e-16   2.32e-15     -0.374      0.708   -5.43e-15
3.69e-15
C(race)[T.21]      -8.4128      5.441     -1.546      0.122     -19.079
2.253
C(race)[T.22]   -1.443e-15    1.9e-15     -0.758      0.449   -5.17e-15
2.29e-15
C(race)[T.23]      -5.2563      7.696     -0.683      0.495     -20.342
```

```
9.829
C(race)[T.24]      1.845e-15   1.33e-15      1.389      0.165   -7.58e-16
4.45e-15
C(race)[T.25]      6.161e-16    1.3e-15      0.474      0.635   -1.93e-15
3.16e-15
C(race)[T.26]     -4.707e-16    1.5e-15     -0.315      0.753    -3.4e-15
2.46e-15
C(ethnic)[T.2.0]      1.4990      0.285      5.262      0.000       0.941
2.057
C(ethnic)[T.3.0]      0.1124      0.424      0.265      0.791      -0.719
0.944
C(ethnic)[T.4.0]     -0.5439      0.432     -1.260      0.208      -1.390
0.302
C(ethnic)[T.5.0]     -0.8408      0.348     -2.417      0.016      -1.523
-0.159
C(ethnic)[T.6.0]     -0.1841      0.301     -0.611      0.541      -0.774
0.406
C(ethnic)[T.7.0]      1.5788      0.332      4.756      0.000       0.928
2.230
C(ethnic)[T.8.0]      1.7910      0.310      5.776      0.000       1.183
2.399
has_college_educ      9.9135      0.488     20.301      0.000       8.956
10.871
age                   0.0843      0.006     14.997      0.000       0.073
0.095
sex                  -2.4613      0.144    -17.039      0.000      -2.745
-2.178
==============================================================================
Omnibus:                     6213.137   Durbin-Watson:                   1.907
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            62966.681
Skew:                           2.349   Prob(JB):                         0.00
Kurtosis:                      13.337   Cond. No.                     3.64e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.48e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
"""
```

Fitting a OLS model using the matched covariates and controlling only for treatment (has college), the mean difference reduces even more to $9.41. We expect that matching parameters, we are reducing the bias in the selection process of the groups.

[78]:
```
# Fitting a weighted least squared model, including only the treatment␣
↪variable, using the weight obtained by propensity score matching
```

```
# Adding matched variables helped ensure we are comparing apples to apples␣
↪dropping the difference in hourly wage $9.41
weight = matched['weight'].values
smf.wls('earnhre_dollars ~ has_college_educ', data = matched, weights = weight).
↪fit().summary()
```

[78]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            WLS Regression Results
==============================================================================
Dep. Variable:        earnhre_dollars   R-squared:                       0.120
Model:                            WLS   Adj. R-squared:                  0.118
Method:                 Least Squares   F-statistic:                     69.51
Date:                Sun, 01 Mar 2020   Prob (F-statistic):           7.05e-16
Time:                        22:08:20   Log-Likelihood:                -2048.5
No. Observations:                 514   AIC:                             4101.
Df Residuals:                     512   BIC:                             4109.
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
====
                       coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Intercept           17.4543      0.841     20.750      0.000      15.802
19.107
has_college_educ     9.4135      1.129      8.337      0.000       7.195
11.632
==============================================================================
Omnibus:                      234.000   Durbin-Watson:                   1.944
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1231.791
Skew:                           1.976   Prob(JB):                    3.31e-268
Kurtosis:                       9.473   Cond. No.                         2.77
==============================================================================

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

Based on the results, accounting for the covariates decreases the coefficient on the treatment variable, we appreciate that as we use a model with balanced sample and demographic covariates (ie : age, race, sex) other than the treatment between treatment and control group, we are able to better interpret the individual effect of having college education on salary, as the mean difference between the control and treatment group is $ 9.79 in the last fitted model.

[79]:
```python
# Fitting the weighted model w/covariants, which increased the college␣
↪education difference to $9.79
smf.wls('earnhre_dollars ~ has_college_educ + age + sex + C(race) + C(ethnic)',␣
↪data = matched, weights = weight).fit().summary()
```

[79]:
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            WLS Regression Results
==============================================================================
Dep. Variable:        earnhre_dollars   R-squared:                       0.199
Model:                            WLS   Adj. R-squared:                  0.168
Method:                 Least Squares   F-statistic:                     6.466
Date:                Sun, 01 Mar 2020   Prob (F-statistic):           4.11e-15
Time:                        22:08:21   Log-Likelihood:                -2024.1
No. Observations:                 514   AIC:                             4088.
Df Residuals:                     494   BIC:                             4173.
Df Model:                          19
Covariance Type:            nonrobust
================================================================================
====
                       coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
----
Intercept           15.8784      3.017      5.263      0.000       9.951
21.806
C(race)[T.2]        -1.0562      2.520     -0.419      0.675      -6.007
3.895
C(race)[T.3]        10.8806      4.007      2.715      0.007       3.007
18.754
C(race)[T.4]         8.3136      3.842      2.164      0.031       0.764
15.863
C(race)[T.5]        -0.6890     11.903     -0.058      0.954     -24.076
22.698
C(race)[T.6]        -1.3422      6.898     -0.195      0.846     -14.896
12.212
C(race)[T.7]        -8.4561      8.360     -1.011      0.312     -24.882
7.970
C(race)[T.8]        -5.8187     11.774     -0.494      0.621     -28.952
17.314
C(race)[T.10]        7.9876     11.873      0.673      0.501     -15.340
31.315
C(race)[T.16]      -12.1882      8.459     -1.441      0.150     -28.809
4.432
C(ethnic)[T.2.0]     5.0951      1.773      2.874      0.004       1.612
8.579
C(ethnic)[T.3.0]    -1.8011      2.613     -0.689      0.491      -6.936
```

```
3.334
C(ethnic)[T.4.0]     2.0434      3.228      0.633      0.527      -4.298
8.385
C(ethnic)[T.5.0]    -3.3632      3.083     -1.091      0.276      -9.420
2.694
C(ethnic)[T.6.0]     1.7360      1.893      0.917      0.359      -1.983
5.455
C(ethnic)[T.7.0]     2.2683      1.716      1.322      0.187      -1.102
5.639
C(ethnic)[T.8.0]     6.5058      1.972      3.299      0.001       2.631
10.381
has_college_educ     9.7862      1.113      8.794      0.000       7.600
11.973
age                  0.0955      0.049      1.946      0.052      -0.001
0.192
sex                 -2.9566      1.137     -2.601      0.010      -5.190
-0.723
==============================================================================
Omnibus:                      220.901   Durbin-Watson:                   2.032
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1146.453
Skew:                           1.848   Prob(JB):                    1.12e-249
Kurtosis:                       9.315   Cond. No.                         962.
==============================================================================

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