PSet1_Q5a_ARE213

October 2, 2023

[1]:
```
import numpy as np
import pandas as pd
import statsmodels.api as sm
from sklearn.linear_model import LogisticRegression
```

[3]:
```
df = pd.read_csv("clean_pset1.csv")

################
#5.a
################

#Logit to estimate propensity score
y_log_reg = df['tobacco']
#outcome
y = ['dbrwt']
#treatment
D = ['tobacco']
#cor with y and D
x1 = ['alcohol', 'mrace3_2', 'mrace3_3', 'ormothhis', 'adeq_2.0', 'adeq_3.0',␣
 ↪'cardiac', 'pre4000', 'phyper',
      'diabetes', 'anemia', 'lung', 'dlivord', 'educ_0.0', 'educ_1.0', 'educ_2.
 ↪0','dmage', 'dmar','tot_2.0',
       'tot_3.0','tot_4.0','tot_5.0','tot_6.0','tot_7.0','tot_8.0','live_1.0',␣
 ↪'live_2.0', 'live_3.0','live_4.0',
       'live_5.0','live_6.0','live_7.0','live_8.0', 'live_9.0']
#cor with D not y
x2 = []
#cor with y not D
x3 = ['dgestat', 'csex', 'plur_1']

X_log_reg = df[x1+x3]

model = LogisticRegression(solver='liblinear', random_state=0)
model.fit(X_log_reg, y_log_reg)

#2nd column gives us predictions
predictions = model.predict_proba(X_log_reg)[:,1]
```

```python
df['predictions'] = model.predict_proba(X_log_reg)[:,1]


#Calculate weights
wt = (df['tobacco'] / predictions) + (1 - df['tobacco']) /(1 -  predictions)

#Demeaned X matrix

for item in x1+x3:
    df[item+'demeaned'] = df[item].sub(df[item].mean())

#multiply by D
for item in x1+x3:
    df['tobacco*'+item] = df[item+'demeaned']*df['tobacco']


#concatenate
d_times_demeaned_X = df[['tobacco*alcohol',
       'tobacco*mrace3_2', 'tobacco*mrace3_3', 'tobacco*ormothhis',
       'tobacco*adeq_2.0', 'tobacco*adeq_3.0', 'tobacco*cardiac',
       'tobacco*pre4000', 'tobacco*phyper',
       'tobacco*diabetes', 'tobacco*anemia', 'tobacco*lung', 'tobacco*dlivord',
       'tobacco*educ_0.0', 'tobacco*educ_1.0', 'tobacco*educ_2.0',
       'tobacco*dgestat', 'tobacco*dmage', 'tobacco*dmar', 'tobacco*csex',
       'tobacco*tot_2.0', 'tobacco*tot_3.0', 'tobacco*tot_4.0',
       'tobacco*tot_5.0', 'tobacco*tot_6.0', 'tobacco*tot_7.0',
       'tobacco*tot_8.0', 'tobacco*live_1.0', 'tobacco*live_2.0',
       'tobacco*live_3.0', 'tobacco*live_4.0', 'tobacco*live_5.0',
       'tobacco*live_6.0', 'tobacco*live_7.0', 'tobacco*live_8.0',
       'tobacco*live_9.0', 'tobacco*plur_1']]

#Outcome: birthweight
y_log_reg = df['dbrwt']

#Create final covariates matrix:
double_robust_reg_X = pd.concat([df['tobacco'],
                                 X_log_reg,
                                 d_times_demeaned_X], axis=1)


fit_wls = sm.WLS(y_log_reg, double_robust_reg_X, weights=wt).fit()
print(fit_wls.summary())
```

```
                            WLS Regression Results
================================================================================
=======
Dep. Variable:                  dbrwt   R-squared (uncentered):
0.981
```

```
Model:                            WLS   Adj. R-squared (uncentered):
0.981
Method:                 Least Squares   F-statistic:
7.849e+04
Date:                Sun, 01 Oct 2023   Prob (F-statistic):
0.00
Time:                        11:46:32   Log-Likelihood:
-8.8421e+05
No. Observations:              114610   AIC:
1.769e+06
Df Residuals:                  114535   BIC:
1.769e+06
Df Model:                          75
Covariance Type:            nonrobust
================================================================================
=====
                       coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-----
tobacco           -207.0867      2.748    -75.358      0.000    -212.473
-201.701
alcohol            -41.5588     20.689     -2.009      0.045     -82.109
-1.008
mrace3_2          -204.0738     13.970    -14.608      0.000    -231.455
-176.692
mrace3_3          -149.2118      6.744    -22.125      0.000    -162.430
-135.994
ormothhis         -106.5636     10.719     -9.942      0.000    -127.573
-85.555
adeq_2.0           -40.6022      4.976     -8.159      0.000     -50.355
-30.849
adeq_3.0           -83.0854      9.988     -8.319      0.000    -102.661
-63.509
cardiac            -25.1022     23.434     -1.071      0.284     -71.032
20.828
pre4000            396.3398     16.385     24.189      0.000     364.226
428.454
phyper            -102.1757     11.280     -9.058      0.000    -124.284
-80.067
diabetes           140.5447     12.034     11.679      0.000     116.959
164.131
anemia              18.2205     19.477      0.936      0.350     -19.954
56.395
lung               -24.6043     22.895     -1.075      0.283     -69.479
20.270
dlivord             25.6250      3.833      6.685      0.000      18.112
33.138
```

educ_0.0	-1796.3924	36.128	-49.723	0.000	-1867.202	-1725.582
educ_1.0	-1811.8350	33.714	-53.742	0.000	-1877.913	-1745.757
educ_2.0	-1781.7611	34.195	-52.107	0.000	-1848.782	-1714.740
dgestat	114.1372	0.793	143.880	0.000	112.582	115.692
dmage	0.6116	0.468	1.308	0.191	-0.305	1.528
dmar	47.1451	5.822	8.098	0.000	35.735	58.555
csex	136.2566	3.873	35.177	0.000	128.665	143.849
tot_2.0	9.3572	7.222	1.296	0.195	-4.799	23.513
tot_3.0	12.6626	8.451	1.498	0.134	-3.902	29.227
tot_4.0	6.5466	10.192	0.642	0.521	-13.430	26.523
tot_5.0	-1.1584	12.878	-0.090	0.928	-26.400	24.083
tot_6.0	-3.1789	16.849	-0.189	0.850	-36.202	29.845
tot_7.0	-19.5211	23.197	-0.842	0.400	-64.987	25.945
tot_8.0	-60.7119	26.373	-2.302	0.021	-112.402	-9.022
live_1.0	-45.1494	26.526	-1.702	0.089	-97.141	6.842
live_2.0	-58.9822	23.476	-2.512	0.012	-104.995	-12.969
live_3.0	48.8789	11.649	4.196	0.000	26.048	71.710
live_4.0	97.5953	10.632	9.179	0.000	76.756	118.434
live_5.0	99.3367	9.437	10.526	0.000	80.840	117.833
live_6.0	91.4001	9.968	9.169	0.000	71.863	110.938
live_7.0	87.3700	11.276	7.748	0.000	65.269	109.471
live_8.0	67.5009	12.922	5.224	0.000	42.174	92.828
live_9.0	56.7237	10.326	5.493	0.000	36.485	76.962
plur_1	557.5078	14.856	37.527	0.000	528.390	586.625

tobacco*alcohol	-37.3531	28.791	-1.297	0.195	-93.784	19.078
tobacco*mrace3_2	168.2648	22.364	7.524	0.000	124.431	212.099
tobacco*mrace3_3	29.6411	9.041	3.279	0.001	11.921	47.361
tobacco*ormothhis	63.4584	15.282	4.152	0.000	33.505	93.411
tobacco*adeq_2.0	-10.4443	6.998	-1.493	0.136	-24.160	3.271
tobacco*adeq_3.0	22.6612	14.154	1.601	0.109	-5.081	50.403
tobacco*cardiac	40.0302	33.516	1.194	0.232	-25.661	105.721
tobacco*pre4000	-81.6672	23.575	-3.464	0.001	-127.875	-35.460
tobacco*phyper	75.7972	15.444	4.908	0.000	45.528	106.066
tobacco*diabetes	108.0868	16.720	6.465	0.000	75.316	140.858
tobacco*anemia	-23.6158	27.305	-0.865	0.387	-77.133	29.901
tobacco*lung	-9.3116	31.898	-0.292	0.770	-71.831	53.208
tobacco*dlivord	-20.9621	5.476	-3.828	0.000	-31.695	-10.229
tobacco*educ_0.0	2020.2018	272.047	7.426	0.000	1486.993	2553.411
tobacco*educ_1.0	2127.1097	271.407	7.837	0.000	1595.157	2659.062
tobacco*educ_2.0	2141.9499	271.485	7.890	0.000	1609.843	2674.057
tobacco*dgestat	-7.9058	1.117	-7.076	0.000	-10.095	-5.716
tobacco*dmage	-2.7495	0.634	-4.337	0.000	-3.992	-1.507
tobacco*dmar	-9.1094	7.804	-1.167	0.243	-24.405	6.186
tobacco*csex	2.9991	5.484	0.547	0.584	-7.749	13.747
tobacco*tot_2.0	11.7417	10.104	1.162	0.245	-8.062	31.545
tobacco*tot_3.0	-10.5351	11.869	-0.888	0.375	-33.799	12.728
tobacco*tot_4.0	41.4315	14.421	2.873	0.004	13.167	69.696
tobacco*tot_5.0	1.9808	18.193	0.109	0.913	-33.677	37.639

```
tobacco*tot_6.0      14.1254     23.972      0.589      0.556     -32.860
61.111
tobacco*tot_7.0      50.9782     32.540      1.567      0.117     -12.800
114.757
tobacco*tot_8.0     116.8786     36.857      3.171      0.002      44.639
189.118
tobacco*live_1.0     17.3771     37.753      0.460      0.645     -56.618
91.372
tobacco*live_2.0      2.8142     32.849      0.086      0.932     -61.569
67.198
tobacco*live_3.0      8.9669     16.619      0.540      0.589     -23.605
41.539
tobacco*live_4.0     -0.7858     15.188     -0.052      0.959     -30.554
28.982
tobacco*live_5.0    -23.2003     13.411     -1.730      0.084     -49.486
3.085
tobacco*live_6.0      0.2926     14.108      0.021      0.983     -27.359
27.944
tobacco*live_7.0     27.2405     16.057      1.696      0.090      -4.232
58.713
tobacco*live_8.0     -0.0084     18.251     -0.000      1.000     -35.780
35.763
tobacco*live_9.0      5.3063     14.470      0.367      0.714     -23.054
33.667
tobacco*plur_1        0.3808     21.088      0.018      0.986     -40.952
41.714
==============================================================================
Omnibus:                    19870.950   Durbin-Watson:                   1.963
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           451511.094
Skew:                           0.126   Prob(JB):                         0.00
Kurtosis:                      12.720   Cond. No.                     1.65e+04
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not
contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[3] The condition number is large, 1.65e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[ ]: