Question (a): Initial Data Pre-processing and Balance Check

1. Define function that transfers nominal variable 'treatment' into dummy variable.

```
def t(x):
    if x == 'Selected':
        x = 1
    else:
        x = 0
    return x
```

```
#transfer treatment into 0-1
data3['treatment']=data2['treatment'].apply(lambda x: t(x))
```

2. Add a column named 'inPortland' to split the data into Portland area and non-Portland area.

```
#inPorland
inPortland=[]
sign = np.array(data3['any_loun_ed'])
for i in range(len(sign)):
    if sign[i]=='Yes':
        inPortland.append(1)
    elif sign[i] == 'No':
        inPortland.append(1)
    else: inPortland.append(0)
data3['inPortland'] = inPortland
```

3. Balance check for Portland area ('inPortland'):

```
features=['treatment','numhh_list_signed self up','numhh_list_signed self up + 1 additional person',
          'numhh_list_signed self up + 2 additional people']
LinearModel_new_spending = sm.OLS(endog = data4[data4['inPortland']==1]['birthyear_list'], exog = sm.add_constant(data4[data4['inPortland']==1][['treatment','numhh
      'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

OLS Regression Results

Dep. Variable:	inPortland	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.002
Method:	Least Squares	F-statistic:	59.02
Date:	Mon, 07 Mar 2022	Prob (F-statistic):	4.25e-38
Time:	12:34:35	Log-Likelihood:	-49627.
No. Observations:	74922	AIC:	9.926e+04
Df Residuals:	74918	BIC:	9.930e+04
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.184e+11	5.78e+11	-0.205	0.838	-1.25e+12	1.01e+12
treatment	-0.0011	0.004	-0.293	0.769	-0.009	0.006
numhh_list_signed self up	1.184e+11	5.78e+11	0.205	0.838	-1.01e+12	1.25e+12
numhh_list_signed self up + 1 additional person	1.184e+11	5.78e+11	0.205	0.838	-1.01e+12	1.25e+12
numhh_list_signed self up + 2 additional people	1.184e+11	5.78e+11	0.205	0.838	-1.01e+12	1.25e+12

Omnibus:	508865.179	Durbin-Watson:	1.993
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13256.075
Skew:	0.725	Prob(JB):	0.00
Kurtosis:	1.536	Cond. No.	9.16e+14

Conclusion: The P>|t| of treatment is 0.769, which is larger than 0.05, implying that the data is balanced.

Y	X	Balance or not
inPortland	Treatment	Balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

4. Balance check for ('birthyear_list'):

```
data4=data3.copy()
```

```
data4[data4['inPortland']==1]['birthyear_list']
```

```
4        1969
7        1968
8        1977
15       1971
17       1957
         ...
74906    1953
74910    1963
74914    1951
74917    1955
74919    1965
Name: birthyear_list, Length: 24646, dtype: int16
```

```
features=['treatment','numhh_list_signed self up','numhh_list_signed self up + 1 additional person',
          'numhh_list_signed self up + 2 additional people']
LinearModel_new_spending = sm.OLS(endog = data4[data4['inPortland']==1]['birthyear_list'], exog = sm.add
      'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         birthyear_list   R-squared:                      -0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    -4.992
Date:                Mon, 07 Mar 2022   Prob (F-statistic):              1.00
Time:                        12:42:54   Log-Likelihood:                -96317.
No. Observations:               24646   AIC:                         1.926e+05
Df Residuals:                   24641   BIC:                         1.927e+05
Df Model:                           4
Covariance Type:            nonrobust
================================================================================================================
                                                   coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------------
const                                       -7.305e+12   1.65e+13     -0.443      0.658   -3.97e+13     2.5e+13
treatment                                       0.1008      0.160      0.628      0.530      -0.214       0.415
numhh_list_signed self up                    7.305e+12   1.65e+13      0.443      0.658    -2.5e+13    3.97e+13
numhh_list_signed self up + 1 additional person  7.305e+12   1.65e+13      0.443      0.658    -2.5e+13    3.97e+13
numhh_list_signed self up + 2 additional people  7.305e+12   1.65e+13      0.443      0.658    -2.5e+13    3.97e+13
==============================================================================
Omnibus:                    12575.280   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            1372.565
Skew:                          -0.119   Prob(JB):                    8.94e-299
Kurtosis:                       1.869   Cond. No.                     5.89e+14
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.33e-25. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Y	X	Balance or not
Birthyear_list	Treatment	balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

Conclusion: The P>|t| of treatment is 0.530, which is larger than 0.05, implying that the data is balanced.

5. Balance check for ‘female_list’ :

```
data5=data4.copy()
```

```
def gender(x):
    if x == '1: Female':
        x = 1
    else:
        x = 0
    return x
```

```
data5['female_list']=data4['female_list'].apply(lambda x: gender(x))
```

```
LinearModel_new_spending = sm.OLS(endog = data5[data5['inPortland']==1]['female_list'], exog = sm.add_constant(data5[data5['inPortland']==1][['treatment','numhh_li
      'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            female_list   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     3.988
Date:                Mon, 07 Mar 2022   Prob (F-statistic):            0.00309
Time:                        13:02:53   Log-Likelihood:                -17772.
No. Observations:               24646   AIC:                         3.555e+04
Df Residuals:                   24641   BIC:                         3.559e+04
Df Model:                           4
Covariance Type:            nonrobust
=============================================================================================================
                                                coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                     -7.117e+11   6.82e+11     -1.044      0.296   -2.05e+12    6.24e+11
treatment                                   -0.0097      0.007     -1.459      0.144      -0.023       0.003
numhh_list_signed self up                 7.117e+11   6.82e+11      1.044      0.296   -6.24e+11    2.05e+12
numhh_list_signed self up + 1 additional person  7.117e+11   6.82e+11      1.044      0.296   -6.24e+11    2.05e+12
numhh_list_signed self up + 2 additional people  7.117e+11   6.82e+11      1.044      0.296   -6.24e+11    2.05e+12
==============================================================================
Omnibus:                    86891.246   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4049.453
Skew:                          -0.187   Prob(JB):                         0.00
Kurtosis:                       1.050   Cond. No.                     5.89e+14
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.33e-25. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Y	X	Balance or not
Female_list	Treatment	Balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

Conclusion: The P>|t| of treatment is 0.144, which is larger than 0.05, implying that the data is balanced.

6. Balance check for 'self_list' :

Transfers 'Signed self up' into dummy variables and then balance check:

```
data6=data5.copy()
```

```
def signself(x):
    if x == 'Signed self up':
        x = 1
    else:
        x=0
    return x
```

```
data6['self_list']=data5['self_list'].apply(lambda x: signself(x))
```

```
LinearModel_new_spending = sm.OLS(endog = data6[data6['inPortland']==1]['self_list'],
                                  exog = sm.add_constant(data6[data6['inPortland']==1][['treatment','numhh_list_signed self up',
    'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              self_list   R-squared:                       0.428
Model:                            OLS   Adj. R-squared:                  0.428
Method:                 Least Squares   F-statistic:                     4603.
Date:                Mon, 07 Mar 2022   Prob (F-statistic):               0.00
Time:                        13:06:32   Log-Likelihood:                 1374.9
No. Observations:               24646   AIC:                            -2740.
Df Residuals:                   24641   BIC:                            -2699.
Df Model:                           4
Covariance Type:            nonrobust
=============================================================================================================
                                                coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                     -9.612e+10   3.13e+11     -0.307      0.759   -7.1e+11    5.18e+11
treatment                                    0.0007      0.003      0.213      0.831      -0.005       0.007
numhh_list_signed self up                 9.612e+10   3.13e+11      0.307      0.759   -5.18e+11    7.1e+11
numhh_list_signed self up + 1 additional person  9.612e+10   3.13e+11      0.307      0.759   -5.18e+11    7.1e+11
numhh_list_signed self up + 2 additional people  9.612e+10   3.13e+11      0.307      0.759   -5.18e+11    7.1e+11
==============================================================================
Omnibus:                     1908.989   Durbin-Watson:                   1.951
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5051.124
Skew:                           0.446   Prob(JB):                         0.00
Kurtosis:                       5.030   Cond. No.                     5.89e+14
==============================================================================
```

Conclusion: The P>|t| of treatment is 0.831, which is larger than 0.05, implying that the data is balanced.

Y	X	Balance or not
Self_list	Treatment	Balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

7. Balance check for 'any_visited_pre_ed' :

Transfer any_visited_pre_ed into dummy variable and then balance check:

Balance check for 'any_visit_pre_ed'

```
data7=data6.copy()
```

```
def yn(x):
    if x=='Yes':
        x=1
    else:
        x=0
    return x
```

```
data7['any_visit_pre_ed']=data6['any_visit_pre_ed'].apply(lambda x: yn(x))
```

```
data7[data7['inPortland']==1]['any_visit_pre_ed'].value_counts()
```

```
0    16930
1     7716
Name: any_visit_pre_ed, dtype: int64
```

```
LinearModel_new_spending = sm.OLS(endog = data7[data7['inPortland']==1]['any_visit_pre_ed'],
                                  exog = sm.add_constant(data7[data7['inPortland']==1][['treatment','numhh_list_signed self up','numhh_list_sign
                                  'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       any_visit_pre_ed   R-squared:                       0.015
Model:                           OLS   Adj. R-squared:                  0.015
Method:                Least Squares   F-statistic:                     92.21
Date:               Mon, 07 Mar 2022   Prob (F-statistic):           5.82e-78
Time:                       13:09:27   Log-Likelihood:                -15850.
No. Observations:              24646   AIC:                         3.171e+04
Df Residuals:                  24641   BIC:                         3.175e+04
Df Model:                          4
Covariance Type:            nonrobust
================================================================================================================================
                                                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------------------
const                                        -7.239e+10    6.3e+11     -0.115      0.909   -1.31e+12    1.16e+12
treatment                                       0.0045      0.006      0.738      0.461      -0.007       0.017
numhh_list_signed self up                    7.239e+10    6.3e+11      0.115      0.909   -1.16e+12    1.31e+12
numhh_list_signed self up + 1 additional person  7.239e+10    6.3e+11      0.115      0.909   -1.16e+12    1.31e+12
numhh_list_signed self up + 2 additional people  7.239e+10    6.3e+11      0.115      0.909   -1.16e+12    1.31e+12
==============================================================================
Omnibus:                   238376.263   Durbin-Watson:                   1.994
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4294.484
Skew:                           0.780   Prob(JB):                         0.00
Kurtosis:                       1.677   Cond. No.                     5.89e+14
==============================================================================
```

Conclusion: The P>|t| of treatment is 0.461, which is larger than 0.05, implying that the data is balanced.

Y	X	Balance or not
Any_visit_pre_ed	Treatment	Balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

8. Balance check for 'num_visit_pre_cens_ed':

```
data8=data7[data7['inPortland']==1][['num_visit_pre_cens_ed',
                                     'treatment',
                                     'numhh_list_signed self up',
                                     'numhh_list_signed self up + 1 additional person',
                                     'numhh_list_signed self up + 2 additional people']].dropna()
```

```
LinearModel_new_spending = sm.OLS(endog = data8['num_visit_pre_cens_ed'],
                                  exog = sm.add_constant(data8[['treatment','numhh_list_signed self up','numhh_list_signed self up + 1 additional person',
                                 'numhh_list_signed self up + 2 additional people']]))
result_new_spending = LinearModel_new_spending.fit()
print(result_new_spending.summary())
```

Result:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     num_visit_pre_cens_ed   R-squared:                       0.010
Model:                               OLS   Adj. R-squared:                  0.009
Method:                    Least Squares   F-statistic:                     59.14
Date:                   Mon, 07 Mar 2022   Prob (F-statistic):           8.92e-50
Time:                           13:12:11   Log-Likelihood:                -50165.
No. Observations:                  24634   AIC:                         1.003e+05
Df Residuals:                      24629   BIC:                         1.004e+05
Df Model:                              4
Covariance Type:               nonrobust
================================================================================================================================
                                                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------------------
const                                       -4.465e+11   3.05e+12     -0.146      0.884   -6.43e+12    5.53e+12
treatment                                       0.0022      0.025      0.090      0.928      -0.046       0.051
numhh_list_signed self up                    4.465e+11   3.05e+12      0.146      0.884   -5.53e+12    6.43e+12
numhh_list_signed self up + 1 additional person  4.465e+11   3.05e+12      0.146      0.884   -5.53e+12    6.43e+12
numhh_list_signed self up + 2 additional people  4.465e+11   3.05e+12      0.146      0.884   -5.53e+12    6.43e+12
==============================================================================
Omnibus:                    22991.233   Durbin-Watson:                   2.018
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           902060.732
Skew:                           4.595   Prob(JB):                         0.00
Kurtosis:                      31.185   Cond. No.                     7.07e+14
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.23e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Conclusion: The P>|t| of treatment is 0.928, which is larger than 0.05, implying that the data is balanced.

Y	X	Balance or not
num_visit_pre_cens_ed	Treatment	Balanced
	Sign self	
	Sign self + 1	
	Sign self +2	

Question (b): Causal effect of being selected by lottery

1. Features selection as follow:

(1). numhh_list (Number of households signed up. To be confirmed.)

(2). birthyear_list (This variable is correlated to age. To be confirmed.)

(3). female_list (Gender should have correlation to the selection. To be confirmed.)

(4). any_visit_ed (If one has been to ED, he/she may be more in need for Medicaid Program. To be confirmed.)

(5). num_visit_cens_ed (If one has been to ED more than others, he/she would be more likely to be in need for Medicaid Program. To be confirmed.)

(6). any_hosp_ed (If one has been to ED and be hospitalized, he/she would be more in need for. Medicaid Program. To be confirmed.)

(7). any_out_ed (Been to ED but outpatient. To be confirmed.)

2. Transfer nominal variables into dummy variables:

```
[29]: _feature = data9[['numhh_list_signed self up','numhh_list_signed self up + 1 additional person','num
um_visit_cens_ed','any_hosp_ed','any_out_ed','ohp_all_ever_firstn_30sep2009','treatment']]
_dummy = pd.get_dummies(df_feature,columns=['any_visit_ed','any_hosp_ed','any_out_ed','ohp_all_ever_
_dummy
```

[29]:

	numhh_list_signed self up	numhh_list_signed self up + 1 additional person	numhh_list_signed self up + 2 additional people	birthyear_list	female_list	num_visit_cens_ed	treatme
4	1	0	0	1969	1	0.0	
7	0	1	0	1968	0	2.0	
8	1	0	0	1977	1	0.0	
15	0	1	0	1971	1	5.0	
17	1	0	0	1957	0	0.0	
...	...	...	...	...	...	...	
74906	1	0	0	1953	0	0.0	
74910	0	1	0	1963	0	0.0	
74914	1	0	0	1951	1	0.0	
74917	1	0	0	1955	0	5.0	
74919	1	0	0	1965	1	0.0	

24622 rows × 11 columns

Balance check:

```
: feature = ['numhh_list_signed self up','numhh_list_signed self up + 1 additional person',
            'numhh_list_signed self up + 2 additional people','birthyear_list','female_list',
            'num_visit_cens_ed',\
             'any_visit_ed_Yes','any_hosp_ed_Yes','any_out_ed_Yes','treatment']
for i in range(len(feature)):
    BalanceCheck = sm.OLS(endog = df_dummy[feature[i]].astype(float), exog = sm.add_constant(df_dumm
    result_BC = BalanceCheck.fit()
    print(result_BC.summary())
```

Result:

X	Y	Balance or not
Being selected by lottery (treatment)	numhh_list_signed self up	imbalanced
	numhh_list_signed self up + 1 additional person	imbalanced
	numhh_list_signed self up + 2 additional people	imbalanced
	birthyear_list	balanced
	female_list	imbalanced
	any_visit_ed	Balanced
	num_visit_cens_ed	Balanced
	any_hosp_ed	Imbalanced
	any_out_ed	Balanced

According to the balance check, 'birthyear_list', 'any_visit_ed', 'num_visit_cens_ed', 'any_out_ed' are balanced. Other features are imbalanced. It is a biased estimation.

Proof of necessity of number of households:

Regress Y (Enrolled into a Medicaid Program) on W (Being selected by lottery) and X.

Proof of number of household:

```
#When we do not delete ['numhh_list_signed self up + 2 additional people']:
BalanceCheck = sm.OLS(endog = df_dummy['ohp_all_ever_firstn_30sep2009_Enrolled'].astype(float), exog = sm.add_constant(df_

result_BC = BalanceCheck.fit()
print(result_BC.summary())
```

```
                                OLS Regression Results
==============================================================================
Dep. Variable:     ohp_all_ever_firstn_30sep2009_Enrolled   R-squared:                       0.112
Model:                                               OLS   Adj. R-squared:                  0.112
Method:                                    Least Squares   F-statistic:                     344.6
Date:                                   Mon, 28 Feb 2022   Prob (F-statistic):               0.00
Time:                                           15:38:22   Log-Likelihood:                -12696.
No. Observations:                                  24622   AIC:                         2.541e+04
Df Residuals:                                      24612   BIC:                         2.549e+04
Df Model:                                              9
Covariance Type:                               nonrobust
=============================================================================================================
                                                coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                         -1.1810      0.319     -3.698      0.000      -1.807      -0.555
numhh_list_signed self up                     -0.3264      0.107     -3.044      0.002      -0.537      -0.116
numhh_list_signed self up + 1 additional person -0.3357    0.107     -3.127      0.002      -0.546      -0.125
numhh_list_signed self up + 2 additional people -0.5189    0.114     -4.532      0.000      -0.743      -0.295
birthyear_list                                 0.0008      0.000      3.678      0.000       0.000       0.001
female_list                                    0.0883      0.005     16.990      0.000       0.078       0.098
num_visit_cens_ed                              0.0086      0.001      5.979      0.000       0.006       0.011
any_visit_ed_Yes                               0.0335      0.022      1.529      0.126      -0.009       0.077
any_hosp_ed_Yes                                0.0594      0.014      4.281      0.000       0.032       0.087
any_out_ed_Yes                                 0.0689      0.022      3.136      0.002       0.026       0.112
treatment                                      0.2445      0.005     45.291      0.000       0.234       0.255
==============================================================================
Omnibus:                     3145.272   Durbin-Watson:                   2.011
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             4154.967
Skew:                           0.984   Prob(JB):                         0.00
Kurtosis:                       2.577   Cond. No.                     9.94e+17
==============================================================================
```

According to the regression, the estimated average treatment effect of being selected by lottery is 0.2445, which means if an individual was selected by lottery, it' s probability of getting enrolled into any Medicaid Program during study period would increase 24.45% on average.

Moreover, there are two reasons why **the number of households should not be deleted:**
Reason 1: From the rules we know that: **Those in a larger household are more likely to be selected into the treatment condition. If we delete the number of household, the causal effect of other variables would be overestimated.
Reason 2: The coefficients of 'numhh_list_signed self up ', 'numhh_list_signed self up + 1 addition person' and 'numhh_list_signed self up + 2 additional people' are respectively -0.3264, -0.3357, -0.5189. The reasonable interpretation of this is: The more households a person sign up, the more possible that s/he did not enroll in any medical program in the period of March 10th, 2008, to Sept 30th, 2009.

Question (c): Causal effect of enrolling into a Medicaid Program on ED Visit.

(i): the causal effect of enrolling into a Medicaid Program on probability of any ED visits during the study period:
1. Features selection:
(1): 'postn_death'(Death post notification date (2008 and 2009): Oregon Vital 74922 Statistics data)
(2): 'birthyear_list'(Birth year: lottery list data)
(3): 'have_phone_list'(Gave a phone number on lottery sign up)
(4): 'english_list'(Individual requested English-language materials)

(5): 'female_list'(Female)

(ii): transfer the nominal variables into dummy variables:

```
data12=pd.get_dummies(data11,columns=['postn_death','have_phone_list','english_list',
                      'female_list','any_visit_ed','any_visit_pre_ed',
                      'ohp_all_ever_firstn_30sep2009','treatment','numhh_list'],drop_first=True).dropna()
```

```
data12
```

	birthyear_list	num_visit_pre_cens_ed	postn_death_Dead	have_phone_list_Gave Phone Number	english_list_Requested English materials	female_list_1: Female	any_visit_ed_Yes	any_visit_pre_ed_Yes	ohp_all_ever_firstn_30sep2009
4	1969	0.0	0	1	0	1	0	0	
7	1968	0.0	0	1	1	0	1	0	
8	1977	1.0	0	1	1	1	0	1	
15	1971	1.0	0	1	1	1	1	1	
17	1957	2.0	0	1	1	0	0	1	
...	...	...	...	...	...	...	...	...	
74906	1953	0.0	0	1	1	0	0	0	
74910	1963	0.0	0	1	1	0	0	0	
74914	1951	0.0	0	1	0	1	0	0	
74917	1955	0.0	0	1	1	0	1	0	
74919	1965	1.0	0	1	1	1	0	1	

24634 rows × 12 columns

Balance check:

Result:

X	Y	Balance or not
Enrolling into a Medicaid. Program.	birthyear_list	imbalanced
	female_list	imbalanced
	postn_death_Dead	imbalanced
	have_phone_list_Gave Phone Number	balanced
	english_list_Requested English materials	balanced
	'numhh_list_signed self up + 1 additional person'	imbalanced
	'numhh_list_signed self up + 2 additional people'	balanced

Calculating estimated ATE using IV2SLS:

```
# Fit a 2sls model.
import linearmodels
from linearmodels.iv import IV2SLS
data12['const']=1
IV_model = IV2SLS(dependent = data12['any_visit_ed_Yes'] ,endog = data12['ohp_all_ever_firstn_30sep2009_Enrolled'],\
                  exog = data12[['birthyear_list','female_list_1: Female','postn_death_Dead','have_phone_list_Gave Phone Number',
                                 'english_list_Requested English materials',
                                 'numhh_list_signed self up + 1 additional person','numhh_list_signed self up + 2 additional people','const']],
                  instruments=data12['treatment_Selected'])
res_2sls = IV_model.fit()
print(res_2sls)
```

```
                          IV-2SLS Estimation Summary
==============================================================================
Dep. Variable:       any_visit_ed_Yes   R-squared:                      0.0512
Estimator:                  IV-2SLS   Adj. R-squared:                 0.0509
No. Observations:             24634   F-statistic:                    1359.3
Date:                Mon, Mar 07 2022   P-value (F-stat)                0.0000
Time:                        19:02:07   Distribution:                  chi2(8)
Cov. Estimator:                robust

                                 Parameter Estimates
============================================================================================================
                                           Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------------------------------------
birthyear_list                               -0.0002     0.0002    -0.8676     0.3856     -0.0007      0.0003
female_list_1: Female                         0.0096     0.0063     1.5119     0.1306     -0.0028      0.0220
postn_death_Dead                              0.3209     0.0307     10.446     0.0000      0.2607      0.3811
have_phone_list_Gave Phone Number             0.0086     0.0088     0.9730     0.3306     -0.0087      0.0258
english_list_Requested English materials      0.1910     0.0074     25.900     0.0000      0.1765      0.2054
numhh_list_signed self up + 1 additional person -0.0928  0.0072    -12.951     0.0000     -0.1068     -0.0788
numhh_list_signed self up + 2 additional people -0.1088  0.0541    -2.0095     0.0445     -0.2149     -0.0027
const                                         0.5824     0.4817     1.2090     0.2267     -0.3618      1.5265
ohp_all_ever_firstn_30sep2009_Enrolled        0.0701     0.0248     2.8250     0.0047      0.0215      0.1187
============================================================================================================

Endogenous: ohp_all_ever_firstn_30sep2009_Enrolled
Instruments: treatment_Selected
Robust Covariance (Heteroskedastic)
Debiased: False
```

By controlling the features, our regression results imply that:

Being enrolled into a Medicaid Program increases the probability that an individual visit Emergency Department at least once. during the study period by **7.01%** on average.
We have 95% confidence that being enrolled into a Medicaid Program will increase the probability that an individual visit Emergency Department at least once during the study period by **(2.15%, 11.87%).**

(ii): the causal effect of enrolling into a Medicaid Program on numbers of ED visits during the study period:
1. Feature selection (basically the same as (i)):
(1): 'postn_death'(Death post notification date (2008 and 2009): Oregon Vital 74922 Statistics data)
(2): 'birthyear_list'(Birth year: lottery list data)
(3): 'have_phone_list'(Gave a phone number on lottery sign up)
(4): 'english_list'(Individual requested english-language materials)
(5): 'female_list'(Female)

2. transfer the nominal variables into dummy variables:

get dummy

```
[39]: data13 = data1.copy()
      data13['inPortland']=inPortland

[40]: data14=data13[data13['inPortland']==1][['postn_death','birthyear_list','have_phone_list','english_list','female_list','any_visit_pre_ed','num_visit_pre_cens_ed','nu

[41]: data15=pd.get_dummies(data14,columns=['postn_death','have_phone_list','english_list','female_list','any_visit_pre_ed','numhh_list','treatment','ohp_all_ever_firstn

[42]: data15

[42]:
```

	birthyear_list	num_visit_pre_cens_ed	num_visit_cens_ed	postn_death_Dead	have_phone_list_Gave Phone Number	english_list_Requested English materials	female_list_1: Female	any_visit_pre_ed_Yes	numhh_list_signed self up + 1 additional person	numhh_l additic
4	1969	0.0	0.0	0	1	0	1	0	0	
7	1968	0.0	2.0	0	1	1	0	0	1	
8	1977	1.0	0.0	0	1	1	1	1	0	
15	1971	1.0	5.0	0	1	1	1	1	1	
17	1957	2.0	0.0	0	1	1	0	1	0	
...	...	...	...	...	...	...	...	...	...	...
74906	1953	0.0	0.0	0	1	1	0	0	0	
74910	1963	0.0	0.0	0	1	1	0	0	1	
74914	1951	0.0	0.0	0	1	0	1	0	0	
74917	1955	0.0	5.0	0	1	1	0	0	0	
74919	1965	1.0	0.0	0	1	1	1	1	0	

24646 rows × 12 columns

3. Balance check:
Result:

X	Y	Balance or not
Enrolling into a Medicaid Program	birthyear_list	imbalanced
	female_list	imbalanced
	postn_death_Dead	imbalanced
	have_phone_list_Gave Phone Number	balanced
	english_list_Requested English materials	balanced
	numhh_list_signed self up + 1 additional person	imbalanced
	numhh_list_signed self up + 2 additional person	balanced

4. Calculating estimated ATE using IV2SLS:

```
# Fit a 2sls model.
import linearmodels
from linearmodels.iv import IV2SLS
data15['const']=1
IV_model = IV2SLS(dependent = data15['num_visit_cens_ed'] ,endog = data15['ohp_all_ever_firstn_30sep2009_Enrolled'],\
                  exog = data15[['birthyear_list', 'num_visit_pre_cens_ed',
       'postn_death_Dead', 'have_phone_list_Gave Phone Number',
       'english_list_Requested English materials', 'female_list_1: Female',
       'any_visit_pre_ed_Yes',
       'numhh_list_signed self up + 1 additional person',
       'numhh_list_signed self up + 2 additional people','const']],
                  instruments=data15['treatment_Selected'])
res_2sls = IV_model.fit()
print(res_2sls)
```

```
                          IV-2SLS Estimation Summary
==============================================================================
Dep. Variable:      num_visit_cens_ed   R-squared:                      0.3422
Estimator:                  IV-2SLS   Adj. R-squared:                 0.3419
No. Observations:             24615   F-statistic:                    3006.8
Date:                Mon, Mar 07 2022   P-value (F-stat)                0.0000
Time:                        19:06:16   Distribution:                chi2(10)
Cov. Estimator:                robust

                                 Parameter Estimates
============================================================================================================
                                                Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------------------------------------
birthyear_list                                 -6.596e-05     0.0010    -0.0684     0.9454     -0.0020      0.0018
num_visit_pre_cens_ed                              0.7548     0.0281     26.849     0.0000      0.6997      0.8099
postn_death_Dead                                   0.7554     0.2022     3.7361     0.0002      0.3591      1.1517
have_phone_list_Gave Phone Number                  0.0649     0.0373     1.7395     0.0819     -0.0082      0.1380
english_list_Requested English materials           0.2975     0.0228     13.075     0.0000      0.2529      0.3421
female_list_1: Female                             -0.0230     0.0269    -0.8550     0.3926     -0.0756      0.0297
any_visit_pre_ed_Yes                              -0.1239     0.0593    -2.0909     0.0365     -0.2401     -0.0078
numhh_list_signed self up + 1 additional person   -0.2031     0.0236    -8.6099     0.0000     -0.2494     -0.1569
numhh_list_signed self up + 2 additional people   -0.2677     0.1381    -1.9386     0.0525     -0.5383      0.0029
const                                              0.2253     1.8965     0.1188     0.9054     -3.4918      3.9424
ohp_all_ever_firstn_30sep2009_Enrolled             0.3788     0.1057     3.5845     0.0003      0.1717      0.5859
============================================================================================================
```

By controlling the features, our regression results imply that:

Being enrolled into a Medicaid Program during study period increases the times that an individual visit Emergency Department during the study period by **0.3788** times on average.

We have 95% confidence that being enrolled into a Medicaid Program will increase the times that an individual visit Emergency Department during the study period by (**0.1717, 0.5859**) times.