

```

import pandas as pd
import statsmodels.formula.api as smf

# Load the dataset
data = pd.read_csv('ps3data.csv')

# Define the treatment variable
data['treatment'] = (data['female'] == 1) & ((data['age'] == 14) | (data['age'] == 15))

# Define the post-treatment period
data['post'] = 1 # Assuming the dataset is after the program was launched

# Initial regression model without additional controls
model_1 = smf.ols('secondaryenroll ~ treatment + post + treatment:post', data=data).fit()

# Print the summary of the first model
print("Model 1 Summary (Without Additional Controls):")
print(model_1.summary())

```

Model 1 Summary (Without Additional Controls):
OLS Regression Results

Dep. Variable:	secondaryenroll	R-squared:	0.007			
Model:	OLS	Adj. R-squared:	0.007			
Method:	Least Squares	F-statistic:	72.25			
Date:	Tue, 21 Nov 2023	Prob (F-statistic):	1.49e-46			
Time:	06:21:40	Log-Likelihood:	-19943.			
No. Observations:	30295	AIC:	3.989e+04			
Df Residuals:	30291	BIC:	3.993e+04			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.197e+10	2.27e+11	0.185	0.854	-4.04e+11	4.87e+11
treatment[T.True]	-9.147e+11	1.22e+12	-0.749	0.454	-3.31e+12	1.48e+12
post	-4.197e+10	2.27e+11	-0.185	0.854	-4.87e+11	4.04e+11
treatment[T.True]:post	9.147e+11	1.22e+12	0.749	0.454	-1.48e+12	3.31e+12
Omnibus:	230607.374	Durbin-Watson:	1.622			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5287.368			
Skew:	0.729	Prob(JB):	0.00			
Kurtosis:	1.564	Cond. No.	9.51e+14			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 7.24e-26. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

```

# Regression model with additional controls for household and village characteristics
model_2 = smf.ols('secondaryenroll ~ treatment + post + treatment:post + hhheadschooll + bpl + media + electricity + mi

# Print the summary of the second model
print("Model 2 Summary (With Additional Controls):")
print(model_2.summary())

```

Model 2 Summary (With Additional Controls):
OLS Regression Results

=====						
Dep. Variable:	secondaryenroll	R-squared:		0.166		
Model:	OLS	Adj. R-squared:		0.166		
Method:	Least Squares	F-statistic:		499.8		
Date:	Tue, 21 Nov 2023	Prob (F-statistic):		0.00		
Time:	06:21:52	Log-Likelihood:		-17208.		
No. Observations:	30148	AIC:		3.444e+04		
Df Residuals:	30135	BIC:		3.455e+04		
Df Model:	12					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	0.0805	0.003	23.358	0.000	0.074	0.087
treatment[T.True]	-0.0446	0.003	-15.803	0.000	-0.050	-0.039
post	0.0805	0.003	23.358	0.000	0.074	0.087
treatment[T.True]:post	-0.0446	0.003	-15.803	0.000	-0.050	-0.039
hhheadschooll	0.0299	0.001	53.273	0.000	0.029	0.031
bpl	-0.0497	0.005	-9.392	0.000	-0.060	-0.039
media	0.1159	0.006	19.954	0.000	0.105	0.127
electricity	0.1179	0.006	18.549	0.000	0.105	0.130
middle	0.0115	0.005	2.153	0.031	0.001	0.022
bank	0.0203	0.010	2.082	0.037	0.001	0.039
postoff	0.0194	0.006	3.043	0.002	0.007	0.032
busdist	-0.0006	0.000	-2.395	0.017	-0.001	-0.000
towndist	0.0003	0.000	1.307	0.191	-0.000	0.001
railwaydist	0.0004	0.000	3.162	0.002	0.000	0.001
hqdist	6.531e-05	0.000	0.527	0.598	-0.000	0.000
=====						
Omnibus:	4515.591	Durbin-Watson:		1.762		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		2775.793		
Skew:	0.618	Prob(JB):		0.00		
Kurtosis:	2.175	Cond. No.		1.55e+17		
=====						

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
 [2] The smallest eigenvalue is 3.49e-27. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

```
# Define and fit the extended regression model with interaction terms
```

```
model_3 = smf.ols('secondaryenroll ~ treatment + post + treatment:post + treatment:longdist + post:longdist + treatment:post:longdist')
```

```
# Print the summary of the extended model
```

```
print("Model 3 Summary (With Distance Interaction):")
```

```
print(model_3.summary())
```

Model 3 Summary (With Distance Interaction):

OLS Regression Results

Dep. Variable:	secondaryenroll	R-squared:	0.167
Model:	OLS	Adj. R-squared:	0.167
Method:	Least Squares	F-statistic:	431.7
Date:	Tue, 21 Nov 2023	Prob (F-statistic):	0.00
Time:	06:22:06	Log-Likelihood:	-17189.
No. Observations:	30148	AIC:	3.441e+04
Df Residuals:	30133	BIC:	3.453e+04
Df Model:	14		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	0.0879	0.004	23.523	0.000	0.081	0.095
treatment[T.True]	-0.0632	0.004	-14.627	0.000	-0.072	-0.055
post	0.0879	0.004	23.523	0.000	0.081	0.095
treatment[T.True]:post	-0.0632	0.004	-14.627	0.000	-0.072	-0.055
treatment[False]:longdist	-0.0246	0.004	-6.105	0.000	-0.032	-0.017
treatment[True]:longdist	0.0203	0.004	4.942	0.000	0.012	0.028
post:longdist	-0.0042	0.003	-1.343	0.179	-0.010	0.002
treatment[T.True]:post:longdist	0.0203	0.004	4.942	0.000	0.012	0.028
hhheadschooll	0.0299	0.001	53.261	0.000	0.029	0.031
bpl	-0.0495	0.005	-9.356	0.000	-0.060	-0.039
media	0.1161	0.006	19.998	0.000	0.105	0.128
electricity	0.1169	0.006	18.370	0.000	0.104	0.129
middle	0.0107	0.005	2.012	0.044	0.000	0.021
bank	0.0176	0.010	1.798	0.072	-0.002	0.037
postoff	0.0194	0.006	3.042	0.002	0.007	0.032
busdist	-0.0006	0.000	-2.051	0.040	-0.001	-2.49e-05
towndist	0.0003	0.000	1.388	0.165	-0.000	0.001
railwaydist	0.0004	0.000	3.204	0.001	0.000	0.001
hqdist	9.023e-05	0.000	0.726	0.468	-0.000	0.000
=====						
Omnibus:	4491.342	Durbin-Watson:	1.762			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2763.128			
Skew:	0.617	Prob(JB):	0.00			
Kurtosis:	2.177	Cond. No.	3.14e+17			
=====						

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is $8.49\text{e-}28$. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.