Problem Set 2: The Determinants of Life Expectancy of the Poor

Harvard University
Spring 2023
Instructor: Gregory Bruich, Ph.D.
Name: Benjamin Zeisberg

- Posted on: 01/30/2023
- Due at: 11:59pm on 02/07/2023

Suggested Imports

In [4]:
```python
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
import statsmodels.api as smf
from stargazer.stargazer import Stargazer
```

Background

The Health Inequality Project uses 1.4 billion observations on income from tax records covering the U.S. population from 1999-2014 to construct income-mortality gradients for each geographic region in the United States. The resulting datasets are publicly available at healthinequality.org.

The map below shows life expectancy at age 40 for men in the bottom quartile of the income distribution for each commuting zone in the United States.

In this problem set, you will use these data to quantify the determinants of life expectancy for these low income men. The extract of the data set, `healthinequality.dta`, is described below and posted on the course website.

Source: The Health Inequality Project (Chetty, Stepner, Abraham, Lin, Scuderi, Turner, Bergeron, and Cutler 2016)

Data Description

File: `healthinequality.dta`

The data consist of $n = 590$ U.S. commuting zones with populations larger than 25,000 in 2000. Commuting zones are geographical aggregations of counties that are similar to metro areas but cover the entire U.S., including rural areas.

For more details on the construction of the variables included in this data set, please see Chetty, Stepner, Abraham, Lin, Scuderi, Turner, Bergeron, and Cutler (2016), which is posted on the course website.

Variable	Definition	Units	Mean
`cz`	Commuting Zone ID	n/a	n/a
`czame`	Commuting Zone Name	n/a	n/a
`stateabbrv`	2-letter state name (U.S. postal code)	n/a	n/a
`fips`	State FIPS code	n/a	n/a
`life_exp`	Male life expectancy at age 40 for the bottom quartile of the national income distribution (race adjusted)	Years	76.41
`cur_smoke`	Fraction of CZ that currently smokes in the bottom quartile of the national income distribution	Decimal, range 0 to 1	0.2792
`bmi_obese`	Fraction of CZ that is obese in the bottom quartile of the national income distribution	Decimal, range 0 to 1	0.3037
`exercise`	Fraction of CZ that exercised in the past 30 days in the bottom quartile of the national income distribution	Decimal, range 0 to 1	0.6047

Data Load

In [16]:
```python
# Read dataset into a pandas dataframe
health = pd.read_stata("healthinequality.dta")

# Display first 5 rows of data
health.head()
```

Out[16]:

	cz	czname	fips	stateabbrv	cur_smoke	bmi_obese	exercise	life_exp
0	100	Johnson City	47	TN	0.351208	0.292475	0.546503	75.968956
1	200	Morristown	47	TN	0.419753	0.302632	0.506173	75.419853
2	301	Middlesborough	47	TN	0.364103	0.320856	0.475309	76.098907
3	302	Knoxville	47	TN	0.324125	0.259366	0.581132	75.089340
4	401	Winston-Salem	37	NC	0.323423	0.285605	0.606607	75.910576

Instructions

Please submit your Problem Set on Canvas. Your submission should include two files:

1. This notebook as a `.ipynb` file with your code and answers to questions
2. A `.pdf` version of this notebook. TODO: Provide general instructions on converting `.ipynb` to `pdf`

Questions

Note: Short answers should be very succinct. Show your work and intuition clearly: credit is given for explanations and not just having the correct answer

1

Use the starter script files to help you get started on this question. The R^2 regression diagnostic statistic measures how much of the variance in the dependent variable can be explained *linearly* by the covariates in the regression. It equals the ratio between the explained sum of squares and the total sum of squares in a regression:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

In a simple bivariate regression, it equals the square of the correlation coefficient between the dependent variable and the single independent variable.

1. Estimate a regression of `life_exp` on `cur_smoke`. Explain in words what the coefficient on `cur_smoke` means.
2. What is the R^2 of this regression? Does the R^2 tell us whether the regression does a good job of fitting the data? If not, what does it tell us?
3. Now generate a random variable that is independent of both `life_exp` and `cur_smoke`. Run a regression of `life_exp` on an intercept, `cur_smoke`, and the random variable that you generated. What happens to the R^2?
4. Now generate a total of 588 random variables that are independent of each other, `life_exp` and `cur_smoke`. Regress `life_exp` on an intercept, `cur_smoke`, and the 588 random variables that you generated. What happens to the R^2? Discuss briefly.
5. Explain (d) using the following simple non-trivial example: with N = 2 observations, what happens if we run a regression with an intercept and one explanatory variable?
6. Now generate a total of 589 random variables. Regress `life_exp` on an intercept, `cur_smoke`, and the 589 independent random variable that you generated. What happens?
7. Use the residual regression formula to explain (f).

In [17]:
```python
# Your Code Here
# 1. Regression of life_exp on cur_smoke
mod_1 = sm.ols("life_exp ~ cur_smoke",data=health)
res_1 = mod_1.fit(cov_type='HC2')

print(res_1.summary(slim=True))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               life_exp   R-squared:                       0.121
Model:                            OLS   Adj. R-squared:                  0.119
No. Observations:                 590   F-statistic:                     43.65
Covariance Type:                  HC2   Prob (F-statistic):           8.81e-11
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     78.4936      0.329    238.304      0.000      77.848      79.139
cur_smoke     -7.4460      1.127     -6.607      0.000      -9.655      -5.237
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC2)
```

Question 1.

We estimate the average effect of the fraction of people in a commuting zone that currently smokes in the bottom quartile of the national income distribution on life expecantcy to be -7.446 years. Our coefficient `cur_smoke` can be explained as the fraction of people in commuting zone *CZ* that currently smokes in the bottom quartile of the income distribution.

Question 2.

What is the R2 of this regression? The R^2 of this regression is 0.121. Does the R2 tell us whether the regression does a good job of fitting the data? the R^2 does generally not tell us if the regression does a good job of fitting the data, we can see that by plotting the estimated regression line vs the data to see if our estimate is a good or bad. If not, what does it tell us? Generelly the R^2 tells us how much of the variance in our dependent variable is explained by our model.

Question 3.

In [22]:
```python
# Find number of rows
def random_variables(x):
    N = health.shape[0]
    # Create column names like rand_5, rand_6...
    random_column_names = [f"rand_{i}" for i in range(x)]

    # Create a new dataframe with just random columns
    random_df = pd.DataFrame(
        np.random.random(size=(N, x)),
        columns=random_column_names
    )

    # Join old and new dataframes
    new_df = pd.concat([health, random_df], axis=1)
    return new_df

frame = random_variables(1)

# 2. Regression of life_exp on cur_smoke, one random variable
mod_2 = sm.ols("life_exp ~ cur_smoke + rand_0",data=frame)
res_2 = mod_2.fit(cov_type='HC2')

print(res_2.summary(slim=True))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               life_exp   R-squared:                       0.121
Model:                            OLS   Adj. R-squared:                  0.118
No. Observations:                 590   F-statistic:                     21.82
Covariance Type:                  HC2   Prob (F-statistic):           7.21e-10
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     78.4493      0.330    237.631      0.000      77.802      79.096
cur_smoke     -7.4469      1.128     -6.601      0.000      -9.658      -5.236
rand_0         0.0900      0.173      0.519      0.604      -0.250       0.430
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC2)
```

What happens to R^2 ?

Nothing. R^2 is a measure of how much of the variance in the dependent variable is explained by our model, since we are simply adding one random variable we would expect that there be not much change to the R^2 besides a minor increase. An increase because we now have a random variable with random data, that could explain some of the variance in the dependent variable.

Question 4.

Now generate a total of 588 random variables. What happens to R^2? we can simply call the helper function as defined before to create 588 random variables, and then run our regression.

Answer: We find that our R^2 now is 1.00 meaning that our model explains the variance in our dependent variable. Why? Essentially, we have created enough random variables with random data, that some are correlated with our dependent variable, and so we explained all the variance in our dependent variable with our model.

In [26]:
```python
frame = random_variables(588)
columns = frame.columns
seperator = "+"
string = seperator.join(columns[8:])
mod_3 = sm.ols(f"life_exp ~ cur_smoke + {string}", data=frame)
res_3 = mod_3.fit(cov_type='HC2')
print(res_3.summary())
```

```
/opt/homebrew/lib/python3.10/site-packages/statsmodels/regression/linear_model.py:1765: RuntimeWarning: divide by zero encountered in divide
  return 1 - (np.divide(self.nobs - self.k_constant, self.df_resid)
/opt/homebrew/lib/python3.10/site-packages/statsmodels/regression/linear_model.py:1765: RuntimeWarning: invalid value encountered in double_scalars
  return 1 - (np.divide(self.nobs - self.k_constant, self.df_resid)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               life_exp   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                    nan
Method:                 Least Squares   F-statistic:                 3.131e+12
Date:                Tue, 07 Feb 2023   Prob (F-statistic):                nan
Time:                        12:06:22   Log-Likelihood:                 15153.
No. Observations:                 590   AIC:                        -2.913e+04
Df Residuals:                       0   BIC:                        -2.654e+04
Df Model:                         589
Covariance Type:                  HC2
==============================================================================
```

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	150.9371	2.57e-05	5.87e+06	0.000	150.937	150.937
cur_smoke	8.8501	nan	nan	nan	nan	nan
rand_0	-7.5130	nan	nan	nan	nan	nan
rand_1	2.9623	nan	nan	nan	nan	nan
rand_2	7.7262	nan	nan	nan	nan	nan
rand_3	-2.9550	nan	nan	nan	nan	nan
rand_4	-0.2513	nan	nan	nan	nan	nan
rand_5	-8.6829	nan	nan	nan	nan	nan
rand_6	-4.1799	2e-06	-2.09e+06	0.000	-4.180	-4.180
rand_7	-0.5874	nan	nan	nan	nan	nan
rand_8	-10.7204	nan	nan	nan	nan	nan
rand_9	-3.0424	nan	nan	nan	nan	nan
rand_10	-9.4660	nan	nan	nan	nan	nan
rand_11	0.4752	nan	nan	nan	nan	nan
rand_12	-2.6850	nan	nan	nan	nan	nan
rand_13	-2.0554	nan	nan	nan	nan	nan
rand_14	-4.2406	nan	nan	nan	nan	nan
rand_15	5.6165	nan	nan	nan	nan	nan
rand_16	2.2385	1.68e-06	1.33e+06	0.000	2.238	2.238
rand_17	4.3512	nan	nan	nan	nan	nan
rand_18	0.3391	2.96e-06	1.14e+05	0.000	0.339	0.339
rand_19	3.1783	nan	nan	nan	nan	nan
rand_20	10.9034	nan	nan	nan	nan	nan
rand_21	-20.4846	nan	nan	nan	nan	nan
rand_22	-9.4769	nan	nan	nan	nan	nan
rand_23	8.2919	nan	nan	nan	nan	nan
rand_24	6.2936	nan	nan	nan	nan	nan
rand_25	4.0665	2.26e-06	1.8e+06	0.000	4.066	4.066
rand_26	2.5067	nan	nan	nan	nan	nan
rand_27	15.9171	nan	nan	nan	nan	nan
rand_28	-9.4447	nan	nan	nan	nan	nan
rand_29	8.6501	nan	nan	nan	nan	nan
rand_30	-1.3123	nan	nan	nan	nan	nan
rand_31	7.5416	nan	nan	nan	nan	nan
rand_32	8.0483	nan	nan	nan	nan	nan
rand_33	0.8987	nan	nan	nan	nan	nan
rand_34	-8.0404	nan	nan	nan	nan	nan
rand_35	-4.1815	nan	nan	nan	nan	nan
rand_36	3.9498	nan	nan	nan	nan	nan
rand_37	15.5916	nan	nan	nan	nan	nan
rand_38	5.2353	nan	nan	nan	nan	nan
rand_39	-3.3249	nan	nan	nan	nan	nan
rand_40	4.0077	nan	nan	nan	nan	nan
rand_41	-2.4492	nan	nan	nan	nan	nan
rand_42	-0.3659	nan	nan	nan	nan	nan
rand_43	0.7587	8.35e-07	9.09e+05	0.000	0.759	0.759
rand_44	-8.4681	nan	nan	nan	nan	nan
rand_45	-2.6439	nan	nan	nan	nan	nan
rand_46	-6.9437	nan	nan	nan	nan	nan
rand_47	1.1000	nan	nan	nan	nan	nan
rand_48	1.0712	nan	nan	nan	nan	nan
rand_49	4.7752	nan	nan	nan	nan	nan
rand_50	3.8653	1.92e-06	2.01e+06	0.000	3.865	3.865
rand_51	9.7533	nan	nan	nan	nan	nan
rand_52	-5.1049	3.89e-06	-1.31e+06	0.000	-5.105	-5.105
rand_53	-8.6758	4.07e-06	-2.13e+06	0.000	-8.676	-8.676
rand_54	1.6934	nan	nan	nan	nan	nan
rand_55	5.4086	nan	nan	nan	nan	nan
rand_56	5.0966	nan	nan	nan	nan	nan
rand_57	3.3064	nan	nan	nan	nan	nan
rand_58	0.4567	1.27e-07	3.59e+06	0.000	0.457	0.457
rand_59	-5.0857	nan	nan	nan	nan	nan
rand_60	6.5360	nan	nan	nan	nan	nan
rand_61	-18.6387	nan	nan	nan	nan	nan
rand_62	0.3691	nan	nan	nan	nan	nan
rand_63	4.1298	nan	nan	nan	nan	nan
rand_64	4.2224	nan	nan	nan	nan	nan
rand_65	-7.0058	nan	nan	nan	nan	nan
rand_66	-8.6620	nan	nan	nan	nan	nan
rand_67	-4.4356	nan	nan	nan	nan	nan
rand_68	4.2368	3.69e-06	1.15e+06	0.000	4.237	4.237
rand_69	2.9744	3.12e-06	9.55e+05	0.000	2.974	2.974
rand_70	-8.2101	nan	nan	nan	nan	nan
rand_71	-13.8221	nan	nan	nan	nan	nan
rand_72	-3.9435	nan	nan	nan	nan	nan
rand_73	6.2031	nan	nan	nan	nan	nan
rand_74	-2.9414	nan	nan	nan	nan	nan
rand_75	5.0756	4.34e-06	1.17e+06	0.000	5.076	5.076
rand_76	-1.1788	2.33e-07	-5.07e+06	0.000	-1.179	-1.179
rand_77	2.0042	nan	nan	nan	nan	nan
rand_78	9.5188	nan	nan	nan	nan	nan
rand_79	14.5562	3.81e-06	3.82e+06	0.000	14.556	14.556
rand_80	3.3815	nan	nan	nan	nan	nan
rand_81	5.6643	nan	nan	nan	nan	nan
rand_82	-8.5171	nan	nan	nan	nan	nan
rand_83	8.3715	nan	nan	nan	nan	nan
rand_84	-1.6888	nan	nan	nan	nan	nan
rand_85	-10.8568	nan	nan	nan	nan	nan
rand_86	-9.7403	nan	nan	nan	nan	nan
rand_87	9.5342	nan	nan	nan	nan	nan
rand_88	11.5941	nan	nan	nan	nan	nan
rand_89	13.1167	nan	nan	nan	nan	nan
rand_90	7.2594	nan	nan	nan	nan	nan
rand_91	3.0043	nan	nan	nan	nan	nan
rand_92	5.2759	nan	nan	nan	nan	nan
rand_93	2.9484	nan	nan	nan	nan	nan
rand_94	-8.4373	nan	nan	nan	nan	nan
rand_95	0.5469	nan	nan	nan	nan	nan
rand_96	2.2691	nan	nan	nan	nan	nan
rand_97	-5.3605	nan	nan	nan	nan	nan
rand_98	-4.9554	nan	nan	nan	nan	nan
rand_99	-10.2926	nan	nan	nan	nan	nan
rand_100	8.0871	nan	nan	nan	nan	nan
rand_101	2.5281	nan	nan	nan	nan	nan
rand_102	11.5092	nan	nan	nan	nan	nan
rand_103	1.6561	nan	nan	nan	nan	nan
rand_104	4.9558	nan	nan	nan	nan	nan
rand_105	20.5129	nan	nan	nan	nan	nan
rand_106	11.4659	nan	nan	nan	nan	nan
rand_107	-13.3549	nan	nan	nan	nan	nan
rand_108	-0.1599	nan	nan	nan	nan	nan
rand_109	0.5272	nan	nan	nan	nan	nan
rand_110	4.9864	2.85e-06	1.75e+06	0.000	4.986	4.986
rand_111	-3.2747	nan	nan	nan	nan	nan
rand_112	4.4333	nan	nan	nan	nan	nan
rand_113	-8.5680	nan	nan	nan	nan	nan
rand_114	-6.5925	nan	nan	nan	nan	nan
rand_115	2.2402	nan	nan	nan	nan	nan
rand_116	4.9792	nan	nan	nan	nan	nan
rand_117	0.9890	3.18e-06	3.11e+05	0.000	0.989	0.989
rand_118	-1.0330	nan	nan	nan	nan	nan
rand_119	-6.9555	nan	nan	nan	nan	nan
rand_120	-10.6312	nan	nan	nan	nan	nan
rand_121	-1.2178	nan	nan	nan	nan	nan
rand_122	6.5143	nan	nan	nan	nan	nan
rand_123	7.2896	nan	nan	nan	nan	nan
rand_124	-12.7355	nan	nan	nan	nan	nan
rand_125	8.4542	nan	nan	nan	nan	nan
rand_126	2.5100	nan	nan	nan	nan	nan
rand_127	6.2107	nan	nan	nan	nan	nan
rand_128	2.2796	nan	nan	nan	nan	nan
rand_129	-3.6944	nan	nan	nan	nan	nan
rand_130	-2.9240	nan	nan	nan	nan	nan
rand_131	-6.1728	nan	nan	nan	nan	nan
rand_132	-3.2507	1.64e-06	-1.99e+06	0.000	-3.251	-3.251
rand_133	1.9553	nan	nan	nan	nan	nan
rand_134	-7.7139	nan	nan	nan	nan	nan
rand_135	-6.8715	nan	nan	nan	nan	nan
rand_136	-5.9142	nan	nan	nan	nan	nan
rand_137	1.5056	nan	nan	nan	nan	nan
rand_138	4.3985	nan	nan	nan	nan	nan
rand_139	-14.4252	nan	nan	nan	nan	nan
rand_140	-9.7862	nan	nan	nan	nan	nan
rand_141	-4.1467	nan	nan	nan	nan	nan
rand_142	8.7684	nan	nan	nan	nan	nan
rand_143	-11.8490	nan	nan	nan	nan	nan
rand_144	-7.0988	nan	nan	nan	nan	nan
rand_145	5.8352	nan	nan	nan	nan	nan
rand_146	3.6383	nan	nan	nan	nan	nan
rand_147	-1.5948	nan	nan	nan	nan	nan
rand_148	-9.2811	nan	nan	nan	nan	nan
rand_149	-6.0557	nan	nan	nan	nan	nan
rand_150	-9.7536	nan	nan	nan	nan	nan
rand_151	2.1116	nan	nan	nan	nan	nan
rand_152	-8.8651	nan	nan	nan	nan	nan
rand_153	0.7678	nan	nan	nan	nan	nan
rand_154	-3.3153	nan	nan	nan	nan	nan
rand_155	-15.8127	nan	nan	nan	nan	nan
rand_156	-3.2139	2.55e-06	-1.26e+06	0.000	-3.214	-3.214
rand_157	7.6446	nan	nan	nan	nan	nan
rand_158	-5.5250	3.03e-06	-1.82e+06	0.000	-5.525	-5.525
rand_159	-6.2106	nan	nan	nan	nan	nan
rand_160	-5.5109	nan	nan	nan	nan	nan
rand_161	-1.0030	nan	nan	nan	nan	nan
rand_162	7.0909	nan	nan	nan	nan	nan
rand_163	-19.1564	nan	nan	nan	nan	nan
rand_164	1.7499	nan	nan	nan	nan	nan
rand_165	-2.8473	nan	nan	nan	nan	nan
rand_166	4.1292	1.29e-06	3.2e+06	0.000	4.129	4.129
rand_167	-0.1135	nan	nan	nan	nan	nan
rand_168	7.3116	nan	nan	nan	nan	nan
rand_169	11.9854	nan	nan	nan	nan	nan
rand_170	-17.7557	nan	nan	nan	nan	nan
rand_171	-7.8039	nan	nan	nan	nan	nan
rand_172	6.4525	nan	nan	nan	nan	nan
rand_173	9.2341	nan	nan	nan	nan	nan
rand_174	8.1612	nan	nan	nan	nan	nan
rand_175	-0.6630	2.2e-07	-3.01e+06	0.000	-0.663	-0.663
rand_176	1.6322	nan	nan	nan	nan	nan
rand_177	3.0872	nan	nan	nan	nan	nan
rand_178	-3.3783	nan	nan	nan	nan	nan
rand_179	-0.8083	nan	nan	nan	nan	nan
rand_180	-11.4945	nan	nan	nan	nan	nan
rand_181	-0.2250	nan	nan	nan	nan	nan
rand_182	-2.9345	nan	nan	nan	nan	nan
rand_183	-20.2887	nan	nan	nan	nan	nan
rand_184	-0.8340	nan	nan	nan	nan	nan
rand_185	-6.3053	9.89e-07	-6.38e+06	0.000	-6.305	-6.305
rand_186	0.1937	1.5e-06	1.29e+05	0.000	0.194	0.194
rand_187	1.5185	nan	nan	nan	nan	nan
rand_188	-11.0145	nan	nan	nan	nan	nan
rand_189	-16.4249	nan	nan	nan	nan	nan
rand_190	13.5880	nan	nan	nan	nan	nan
rand_191	-10.3184	nan	nan	nan	nan	nan
rand_192	14.4101	nan	nan	nan	nan	nan
rand_193	-20.1427	nan	nan	nan	nan	nan
rand_194	-3.0611	nan	nan	nan	nan	nan
rand_195	-3.3803	nan	nan	nan	nan	nan
rand_196	-0.4028	nan	nan	nan	nan	nan
rand_197	6.5038	nan	nan	nan	nan	nan
rand_198	4.4129	nan	nan	nan	nan	nan
rand_199	7.4496	nan	nan	nan	nan	nan
rand_200	0.7925	1.15e-06	6.89e+05	0.000	0.793	0.793
rand_201	6.5159	nan	nan	nan	nan	nan
rand_202	-3.3103	1.45e-06	-2.29e+06	0.000	-3.310	-3.310
rand_203	3.4782	nan	nan	nan	nan	nan
rand_204	6.8206	nan	nan	nan	nan	nan
rand_205	-7.2895	nan	nan	nan	nan	nan
rand_206	3.7219	nan	nan	nan	nan	nan
rand_207	-16.3152	nan	nan	nan	nan	nan
rand_208	6.0570	nan	nan	nan	nan	nan
rand_209	-7.2335	nan	nan	nan	nan	nan
rand_210	-5.8868	nan	nan	nan	nan	nan
rand_211	14.9118	nan	nan	nan	nan	nan
rand_212	3.3985	nan	nan	nan	nan	nan
rand_213	3.0740	nan	nan	nan	nan	nan
rand_214	0.5758	nan	nan	nan	nan	nan
rand_215	1.5849	1.26e-06	1.26e+06	0.000	1.585	1.585
rand_216	7.5308	nan	nan	nan	nan	nan
rand_217	-2.8927	2.45e-06	-1.18e+06	0.000	-2.893	-2.893
rand_218	7.7519	nan	nan	nan	nan	nan
rand_219	-8.0400	nan	nan	nan	nan	nan
rand_220	-3.5370	nan	nan	nan	nan	nan
rand_221	4.0953	nan	nan	nan	nan	nan
rand_222	-1.7312	nan	nan	nan	nan	nan
rand_223	14.9475	nan	nan	nan	nan	nan
rand_224	-19.9471	nan	nan	nan	nan	nan
rand_225	3.9683	nan	nan	nan	nan	nan
rand_226	6.7189	nan	nan	nan	nan	nan
rand_227	-1.6022	nan	nan	nan	nan	nan
rand_228	1.0397	nan	nan	nan	nan	nan
rand_229	0.4001	nan	nan	nan	nan	nan
rand_230	1.7363	9.12e-07	1.9e+06	0.000	1.736	1.736
rand_231	-4.3588	nan	nan	nan	nan	nan
rand_232	-1.6429	nan	nan	nan	nan	nan
rand_233	-5.2291	nan	nan	nan	nan	nan
rand_234	13.6568	nan	nan	nan	nan	nan
rand_235	-3.2487	nan	nan	nan	nan	nan
rand_236	-7.3592	nan	nan	nan	nan	nan
rand_237	1.7710	nan	nan	nan	nan	nan
rand_238	-14.4033	nan	nan	nan	nan	nan
rand_239	-7.8128	nan	nan	nan	nan	nan
rand_240	-5.6180	nan	nan	nan	nan	nan
rand_241	7.1594	nan	nan	nan	nan	nan
rand_242	-10.2215	nan	nan	nan	nan	nan
rand_243	7.0266	2.19e-06	3.21e+06	0.000	7.027	7.027
rand_244	-2.1151	nan	nan	nan	nan	nan
rand_245	6.3002	nan	nan	nan	nan	nan
rand_246	-0.9871	nan	nan	nan	nan	nan
rand_247	-12.4129	nan	nan	nan	nan	nan
rand_248	-1.7299	nan	nan	nan	nan	nan
rand_249	2.5616	nan	nan	nan	nan	nan
rand_250	11.9870	nan	nan	nan	nan	nan
rand_251	2.9831	nan	nan	nan	nan	nan
rand_252	-4.8682	5.2e-07	-9.36e+06	0.000	-4.868	-4.868
rand_253	-7.6507	nan	nan	nan	nan	nan
rand_254	4.3107	2.19e-06	1.97e+06	0.000	4.311	4.311
rand_255	8.4399	nan	nan	nan	nan	nan
rand_256	-4.2042	nan	nan	nan	nan	nan
rand_257	6.3947	nan	nan	nan	nan	nan
rand_258	-1.9353	nan	nan	nan	nan	nan
rand_259	-3.3031	nan	nan	nan	nan	nan
rand_260	-12.5793	nan	nan	nan	nan	nan
rand_261	-8.8272	nan	nan	nan	nan	nan
rand_262	-7.0562	nan	nan	nan	nan	nan
rand_263	6.3421	nan	nan	nan	nan	nan
rand_264	6.9000	nan	nan	nan	nan	nan
rand_265	-6.9922	nan	nan	nan	nan	nan
rand_266	9.8304	nan	nan	nan	nan	nan
rand_267	-7.4139	nan	nan	nan	nan	nan
rand_268	18.2526	nan	nan	nan	nan	nan
rand_269	0.3932	nan	nan	nan	nan	nan
rand_270	1.0053	nan	nan	nan	nan	nan
rand_271	17.5535	nan	nan	nan	nan	nan
rand_272	7.4271	nan	nan	nan	nan	nan
rand_273	3.4910	3.85e-06	9.07e+05	0.000	3.491	3.491
rand_274	-14.2395	nan	nan	nan	nan	nan
rand_275	1.8030	nan	nan	nan	nan	nan
rand_276	-19.9995	nan	nan	nan	nan	nan
rand_277	8.3291	nan	nan	nan	nan	nan
rand_278	11.7725	nan	nan	nan	nan	nan
rand_279	2.5017	nan	nan	nan	nan	nan
rand_280	4.3654	nan	nan	nan	nan	nan
rand_281	-0.0765	nan	nan	nan	nan	nan
rand_282	2.0409	nan	nan	nan	nan	nan
rand_283	6.2707	nan	nan	nan	nan	nan
rand_284	-7.7954	nan	nan	nan	nan	nan
rand_285	7.5732	nan	nan	nan	nan	nan
rand_286	4.8815	nan	nan	nan	nan	nan
rand_287	6.7448	nan	nan	nan	nan	nan
rand_288	4.9974	nan	nan	nan	nan	nan
rand_289	-8.9067	nan	nan	nan	nan	nan
rand_290	-8.5339	2.92e-06	-2.93e+06	0.000	-8.534	-8.534
rand_291	-7.4248	nan	nan	nan	nan	nan
rand_292	-5.2281	nan	nan	nan	nan	nan
rand_293	-1.1779	nan	nan	nan	nan	nan
rand_294	-2.7730	nan	nan	nan	nan	nan
rand_295	-7.3959	nan	nan	nan	nan	nan
rand_296	-3.8026	nan	nan	nan	nan	nan
rand_297	-4.4013	nan	nan	nan	nan	nan
rand_298	-2.2530	nan	nan	nan	nan	nan
rand_299	-10.6933	nan	nan	nan	nan	nan
rand_300	0.2529	nan	nan	nan	nan	nan
rand_301	0.8838	nan	nan	nan	nan	nan
rand_302	5.8973	5.32e-06	1.11e+06	0.000	5.897	5.897
rand_303	-2.7278	nan	nan	nan	nan	nan
rand_304	2.0452	nan	nan	nan	nan	nan
rand_305	-2.8629	1.57e-06	-1.82e+06	0.000	-2.863	-2.863
rand_306	5.4846	4.16e-06	1.32e+06	0.000	5.485	5.485
rand_307	5.0904	nan	nan	nan	nan	nan
rand_308	5.8851	nan	nan	nan	nan	nan
rand_309	-1.9813	nan	nan	nan	nan	nan
rand_310	-0.7937	nan	nan	nan	nan	nan
rand_311	4.2410	nan	nan	nan	nan	nan
rand_312	5.2977	nan	nan	nan	nan	nan
rand_313	-1.2673	nan	nan	nan	nan	nan
rand_314	10.7175	nan	nan	nan	nan	nan
rand_315	3.2175	nan	nan	nan	nan	nan
rand_316	1.7571	nan	nan	nan	nan	nan
rand_317	4.1252	2e-06	2.06e+06	0.000	4.125	4.125
rand_318	5.9567	nan	nan	nan	nan	nan
rand_319	5.3140	nan	nan	nan	nan	nan
rand_320	12.4683	nan	nan	nan	nan	nan
rand_321	-1.2216	1.35e-06	-9.08e+05	0.000	-1.222	-1.222
rand_322	-15.0232	nan	nan	nan	nan	nan
rand_323	3.7435	nan	nan	nan	nan	nan
rand_324	0.9845	nan	nan	nan	nan	nan
rand_325	-3.1417	nan	nan	nan	nan	nan
rand_326	-0.5533	nan	nan	nan	nan	nan
rand_327	-3.8009	nan	nan	nan	nan	nan
rand_328	7.0277	nan	nan	nan	nan	nan
rand_329	-8.7576	nan	nan	nan	nan	nan
rand_330	-11.7662	nan	nan	nan	nan	nan
rand_331	-5.0856	nan	nan	nan	nan	nan
rand_332	-7.7986	nan	nan	nan	nan	nan
rand_333	2.4134	nan	nan	nan	nan	nan
rand_334	-16.6430	nan	nan	nan	nan	nan
rand_335	2.8897	nan	nan	nan	nan	nan
rand_336	2.7454	nan	nan	nan	nan	nan
rand_337	-7.6572	nan	nan	nan	nan	nan
rand_338	6.7120	nan	nan	nan	nan	nan
rand_339	-10.4271	nan	nan	nan	nan	nan
rand_340	9.6457	nan	nan	nan	nan	nan
rand_341	-8.4014	nan	nan	nan	nan	nan
rand_342	-0.2483	2.52e-06	-9.87e+04	0.000	-0.248	-0.248
rand_343	14.3325	nan	nan	nan	nan	nan
rand_344	1.5331	1.89e-06	8.12e+05	0.000	1.533	1.533
rand_345	-3.5851	nan	nan	nan	nan	nan
rand_346	4.2591	nan	nan	nan	nan	nan
rand_347	-8.0308	nan	nan	nan	nan	nan
rand_348	10.1382	nan	nan	nan	nan	nan
rand_349	-5.5186	nan	nan	nan	nan	nan
rand_350	-8.1762	nan	nan	nan	nan	nan
rand_351	2.7324	1.79e-06	1.52e+06	0.000	2.732	2.732
rand_352	0.2900	nan	nan	nan	nan	nan
rand_353	-6.3722	2.19e-06	-2.91e+06	0.000	-6.372	-6.372
rand_354	-12.9248	nan	nan	nan	nan	nan
rand_355	7.2806	nan	nan	nan	nan	nan
rand_356	2.5661	nan	nan	nan	nan	nan
rand_357	-2.2467	nan	nan	nan	nan	nan
rand_358	4.9770	nan	nan	nan	nan	nan
rand_359	-3.3344	1.66e-06	-2e+06	0.000	-3.334	-3.334
rand_360	-7.1693	nan	nan	nan	nan	nan
rand_361	7.0914	nan	nan	nan	nan	nan
rand_362	4.9958	nan	nan	nan	nan	nan
rand_363	2.5802	1.33e-06	1.95e+06	0.000	2.580	2.580
rand_364	5.2243	2.05e-06	2.54e+06	0.000	5.224	5.224
rand_365	-6.2413	4.85e-06	-1.29e+06	0.000	-6.241	-6.241
rand_366	-1.0349	nan	nan	nan	nan	nan
rand_367	6.4742	nan	nan	nan	nan	nan
rand_368	1.8838	nan	nan	nan	nan	nan
rand_369	13.5118	nan	nan	nan	nan	nan
rand_370	10.3564	nan	nan	nan	nan	nan
rand_371	4.9263	nan	nan	nan	nan	nan
rand_372	-1.9326	nan	nan	nan	nan	nan
rand_373	-8.3864	nan	nan	nan	nan	nan
rand_374	-7.1383	nan	nan	nan	nan	nan
rand_375	-5.2597	nan	nan	nan	nan	nan
rand_376	4.0283	nan	nan	nan	nan	nan
rand_377	6.7231	2.19e-06	3.06e+06	0.000	6.723	6.723
rand_378	-4.6056	3.14e-06	-1.47e+06	0.000	-4.606	-4.606
rand_379	2.6218	nan	nan	nan	nan	nan
rand_380	-3.2535	nan	nan	nan	nan	nan
rand_381	-13.5466	nan	nan	nan	nan	nan
rand_382	-0.7779	3.7e-06	-2.1e+05	0.000	-0.778	-0.778
rand_383	11.1783	nan	nan	nan	nan	nan
rand_384	19.7349	nan	nan	nan	nan	nan
rand_385	-2.1360	nan	nan	nan	nan	nan
rand_386	-5.0561	nan	nan	nan	nan	nan
rand_387	1.9617	1.29e-06	1.52e+06	0.000	1.962	1.962
rand_388	2.3533	1.22e-06	1.93e+06	0.000	2.353	2.353
rand_389	3.8279	nan	nan	nan	nan	nan
rand_390	6.4962	nan	nan	nan	nan	nan
rand_391	1.3610	nan	nan	nan	nan	nan
rand_392	3.0827	nan	nan	nan	nan	nan
rand_393	-5.1971	nan	nan	nan	nan	nan
rand_394	9.9892	nan	nan	nan	nan	nan
rand_395	2.5181	nan	nan	nan	nan	nan
rand_396	-16.7114	nan	nan	nan	nan	nan
rand_397	0.7473	nan	nan	nan	nan	nan
rand_398	-6.8249	nan	nan	nan	nan	nan
rand_399	-16.0769	nan	nan	nan	nan	nan
rand_400	12.1140	nan	nan	nan	nan	nan
rand_401	-6.1223	nan	nan	nan	nan	nan
rand_402	-0.6930	nan	nan	nan	nan	nan
rand_403	-1.5593	nan	nan	nan	nan	nan
rand_404	9.4990	nan	nan	nan	nan	nan
rand_405	-17.1613	nan	nan	nan	nan	nan
rand_406	7.9552	nan	nan	nan	nan	nan
rand_407	2.1450	nan	nan	nan	nan	nan
rand_408	-6.3439	nan	nan	nan	nan	nan
rand_409	1.8111	nan	nan	nan	nan	nan
rand_410	-15.2607	nan	nan	nan	nan	nan
rand_411	6.7350	nan	nan	nan	nan	nan
rand_412	21.2132	nan	nan	nan	nan	nan
rand_413	-9.7045	nan	nan	nan	nan	nan
rand_414	-5.6365	nan	nan	nan	nan	nan
rand_415	18.2317	nan	nan	nan	nan	nan
rand_416	-14.8507	nan	nan	nan	nan	nan
rand_417	11.6989	nan	nan	nan	nan	nan
rand_418	-2.2839	nan	nan	nan	nan	nan
rand_419	-2.6700	nan	nan	nan	nan	nan
rand_420	13.0147	nan	nan	nan	nan	nan
rand_421	-12.8139	nan	nan	nan	nan	nan
rand_422	-6.3955	nan	nan	nan	nan	nan
rand_423	-3.7855	nan	nan	nan	nan	nan
rand_424	-8.3655	nan	nan	nan	nan	nan
rand_425	-11.3365	nan	nan	nan	nan	nan
rand_426	-8.2728	nan	nan	nan	nan	nan
rand_427	-1.9320	nan	nan	nan	nan	nan
rand_428	-0.9021	3.74e-06	-2.41e+05	0.000	-0.902	-0.902
rand_429	-5.0949	nan	nan	nan	nan	nan
rand_430	-9.4831	nan	nan	nan	nan	nan
rand_431	-5.6713	nan	nan	nan	nan	nan
rand_432	7.9250	nan	nan	nan	nan	nan
rand_433	-2.2713	nan	nan	nan	nan	nan
rand_434	0.9876	nan	nan	nan	nan	nan
rand_435	19.5274	nan	nan	nan	nan	nan
rand_436	4.5302	nan	nan	nan	nan	nan
rand_437	-10.3367	nan	nan	nan	nan	nan
rand_438	-6.7344	nan	nan	nan	nan	nan
rand_439	0.8017	1.76e-06	4.54e+05	0.000	0.802	0.802
rand_440	3.2596	nan	nan	nan	nan	nan
rand_441	-3.8374	nan	nan	nan	nan	nan
rand_442	-8.7555	nan	nan	nan	nan	nan
rand_443	2.4163	nan	nan	nan	nan	nan
rand_444	-13.0575	nan	nan	nan	nan	nan
rand_445	6.9306	nan	nan	nan	nan	nan
rand_446	0.0386	nan	nan	nan	nan	nan
rand_447	-4.2560	2.21e-06	-1.92e+06	0.000	-4.256	-4.256
rand_448	-9.5324	nan	nan	nan	nan	nan
rand_449	1.2392	nan	nan	nan	nan	nan
rand_450	15.9637	nan	nan	nan	nan	nan
rand_451	2.0515	nan	nan	nan	nan	nan
rand_452	7.8405	nan	nan	nan	nan	nan
rand_453	-5.9350	nan	nan	nan	nan	nan
rand_454	1.5551	nan	nan	nan	nan	nan
rand_455	-4.4386	nan	nan	nan	nan	nan
rand_456	5.1982	nan	nan	nan	nan	nan
rand_457	2.1967	nan	nan	nan	nan	nan
rand_458	2.3205	nan	nan	nan	nan	nan
rand_459	-3.0858	nan	nan	nan	nan	nan
rand_460	11.1386	nan	nan	nan	nan	nan
rand_461	-17.0373	nan	nan	nan	nan	nan
rand_462	6.0370	nan	nan	nan	nan	nan
rand_463	-7.3365	3.61e-06	-2.03e+06	0.000	-7.337	-7.336
rand_464	-4.5952	nan	nan	nan	nan	nan
rand_465	-10.4497	nan	nan	nan	nan	nan
rand_466	-1.2079	2.97e-06	-4.07e+05	0.000	-1.208	-1.208
rand_467	15.5958	nan	nan	nan	nan	nan
rand_468	5.2120	nan	nan	nan	nan	nan
rand_469	2.7965	nan	nan	nan	nan	nan
rand_470	0.7188	nan	nan	nan	nan	nan
rand_471	-5.0258	nan	nan	nan	nan	nan
rand_472	1.3042	nan	nan	nan	nan	nan
rand_473	-0.0276	nan	nan	nan	nan	nan
rand_474	-6.8044	nan	nan	nan	nan	nan
rand_475	4.7313	nan	nan	nan	nan	nan
rand_476	17.0296	nan	nan	nan	nan	nan
rand_477	10.6133	nan	nan	nan	nan	nan
rand_478	-1.8540	nan	nan	nan	nan	nan
rand_479	-3.8734	nan	nan	nan	nan	nan
rand_480	7.8842	nan	nan	nan	nan	nan
rand_481	-0.7854	1.09e-06	-7.21e+05	0.000	-0.785	-0.785
rand_482	-8.7646	nan	nan	nan	nan	nan
rand_483	-7.6541	nan	nan	nan	nan	nan
rand_484	-7.4450	4.01e-06	-1.86e+06	0.000	-7.445	-7.445
rand_485	-18.1822	nan	nan	nan	nan	nan
rand_486	4.1115	nan	nan	nan	nan	nan
rand_487	-1.1738	nan	nan	nan	nan	nan
rand_488	6.4787	nan	nan	nan	nan	nan
rand_489	-3.1505	nan	nan	nan	nan	nan
rand_490	-1.7128	3.59e-06	-4.77e+05	0.000	-1.713	-1.713
rand_491	-11.5470	nan	nan	nan	nan	nan
rand_492	-3.4318	nan	nan	nan	nan	nan
rand_493	-5.0788	nan	nan	nan	nan	nan
rand_494	-5.5338	nan	nan	nan	nan	nan
rand_495	1.7214	1.22e-06	1.41e+06	0.000	1.721	1.721
rand_496	5.1816	nan	nan	nan	nan	nan
rand_497	-9.2539	1.38e-06	-6.7e+06	0.000	-9.254	-9.254
rand_498	-11.8194	nan	nan	nan	nan	nan
rand_499	-2.4974	nan	nan	nan	nan	nan
rand_500	-1.7589	nan	nan	nan	nan	nan
rand_501	-4.1246	nan	nan	nan	nan	nan
rand_502	-4.5620	nan	nan	nan	nan	nan
rand_503	-1.0998	nan	nan	nan	nan	nan
rand_504	-0.1244	8.5e-07	-1.46e+05	0.000	-0.124	-0.124
rand_505	18.8834	nan	nan	nan	nan	nan
rand_506	15.6888	nan	nan	nan	nan	nan
rand_507	-9.5375	nan	nan	nan	nan	nan
rand_508	-1.0843	nan	nan	nan	nan	nan
rand_509	-5.4462	nan	nan	nan	nan	nan
rand_510	4.0670	nan	nan	nan	nan	nan
rand_511	2.7055	9.53e-07	2.84e+06	0.000	2.706	2.706
rand_512	10.4389	nan	nan	nan	nan	nan
rand_513	8.4147	nan	nan	nan	nan	nan
rand_514	-3.0428	nan	nan	nan	nan	nan
rand_515	2.5308	nan	nan	nan	nan	nan
rand_516	0.6863	nan	nan	nan	nan	nan
rand_517	-6.4093	nan	nan	nan	nan	nan
rand_518	2.5120	nan	nan	nan	nan	nan
rand_519	4.9896	nan	nan	nan	nan	nan
rand_520	-7.4515	nan	nan	nan	nan	nan
rand_521	7.2938	2.95e-06	2.48e+06	0.000	7.294	7.294
rand_522	-1.3917	nan	nan	nan	nan	nan
rand_523	-2.2565	nan	nan	nan	nan	nan
rand_524	-5.4146	nan	nan	nan	nan	nan
rand_525	3.7887	2.39e-06	1.58e+06	0.000	3.789	3.789
rand_526	-9.0916	nan	nan	nan	nan	nan
rand_527	-10.4944	nan	nan	nan	nan	nan
rand_528	-4.4795	nan	nan	nan	nan	nan
rand_529	14.0999	nan	nan	nan	nan	nan
rand_530	-6.4237	3.18e-06	-2.02e+06	0.000	-6.424	-6.424
rand_531	-8.7111	nan	nan	nan	nan	nan
rand_532	-11.4794	nan	nan	nan	nan	nan
rand_533	-11.7979	nan	nan	nan	nan	nan
rand_534	6.4328	nan	nan	nan	nan	nan
rand_535	21.3231	nan	nan	nan	nan	nan
rand_536	18.2962	nan	nan	nan	nan	nan
rand_537	-12.7250	nan	nan	nan	nan	nan
rand_538	7.6351	nan	nan	nan	nan	nan
rand_539	5.4021	nan	nan	nan	nan	nan
rand_540	-15.0423	nan	nan	nan	nan	nan
rand_541	-8.9673	nan	nan	nan	nan	nan
rand_542	11.0829	nan	nan	nan	nan	nan
rand_543	-10.3880	nan	nan	nan	nan	nan
rand_544	7.6674	nan	nan	nan	nan	nan
rand_545	-4.9881	nan	nan	nan	nan	nan
rand_546	-5.0526	nan	nan	nan	nan	nan
rand_547	0.7939	nan	nan	nan	nan	nan
rand_548	6.0857	nan	nan	nan	nan	nan
rand_549	-4.1520	nan	nan	nan	nan	nan
rand_550	5.6915	nan	nan	nan	nan	nan
rand_551	-3.7219	nan	nan	nan	nan	nan
rand_552	-10.1153	nan	nan	nan	nan	nan
rand_553	0.1811	1.38e-06	1.31e+05	0.000	0.181	0.181
rand_554	2.0955	nan	nan	nan	nan	nan
rand_555	17.4720	nan	nan	nan	nan	nan
rand_556	-4.6690	nan	nan	nan	nan	nan
rand_557	6.8384	nan	nan	nan	nan	nan
rand_558	9.8616	nan	nan	nan	nan	nan
rand_559	1.4620	nan	nan	nan	nan	nan
rand_560	12.3204	nan	nan	nan	nan	nan
rand_561	-3.5453	1.06e-06	-3.36e+06	0.000	-3.545	-3.545
rand_562	-20.8977	nan	nan	nan	nan	nan
rand_563	5.1056	2.32e-06	2.2e+06	0.000	5.106	5.106
rand_564	0.2573	nan	nan	nan	nan	nan
rand_565	4.1806	nan	nan	nan	nan	nan
rand_566	9.1501	nan	nan	nan	nan	nan
rand_567	4.9765	nan	nan	nan	nan	nan
rand_568	14.8681	nan	nan	nan	nan	nan
rand_569	-4.0118	nan	nan	nan	nan	nan
rand_570	-7.1489	nan	nan	nan	nan	nan
rand_571	10.1455	nan	nan	nan	nan	nan
rand_572	14.7980	nan	nan	nan	nan	nan
rand_573	-0.7783	nan	nan	nan	nan	nan
rand_574	-10.3130	nan	nan	nan	nan	nan
rand_575	-7.7297	nan	nan	nan	nan	nan
rand_576	6.9833	nan	nan	nan	nan	nan
rand_577	-1.6372	nan	nan	nan	nan	nan
rand_578	-8.8796	nan	nan	nan	nan	nan
rand_579	0.1718	nan	nan	nan	nan	nan
rand_580	2.6320	1.43e-06	1.85e+06	0.000	2.632	2.632
rand_581	11.7525	nan	nan	nan	nan	nan
rand_582	3.9290	nan	nan	nan	nan	nan
rand_583	-4.7138	nan	nan	nan	nan	nan
rand_584	6.2381	nan	nan	nan	nan	nan
rand_585	-11.4354	nan	nan	nan	nan	nan
rand_586	1.8826	nan	nan	nan	nan	nan
rand_587	-10.6667	nan	nan	nan	nan	nan

```
==============================================================================
Omnibus:                        7.647   Durbin-Watson:                   0.242
Prob(Omnibus):                  0.022   Jarque-Bera (JB):               10.895
Skew:                          -0.072   Prob(JB):                      0.00431
Kurtosis:                       3.650   Cond. No.                     2.35e+04
==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC2)
[2] The condition number is large, 2.35e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
/opt/homebrew/lib/python3.10/site-packages/statsmodels/regression/linear_model.py:1854: RuntimeWarning: invalid value encountered in sqrt
  return np.sqrt(np.diag(self.cov_params()))
```

In [10]:

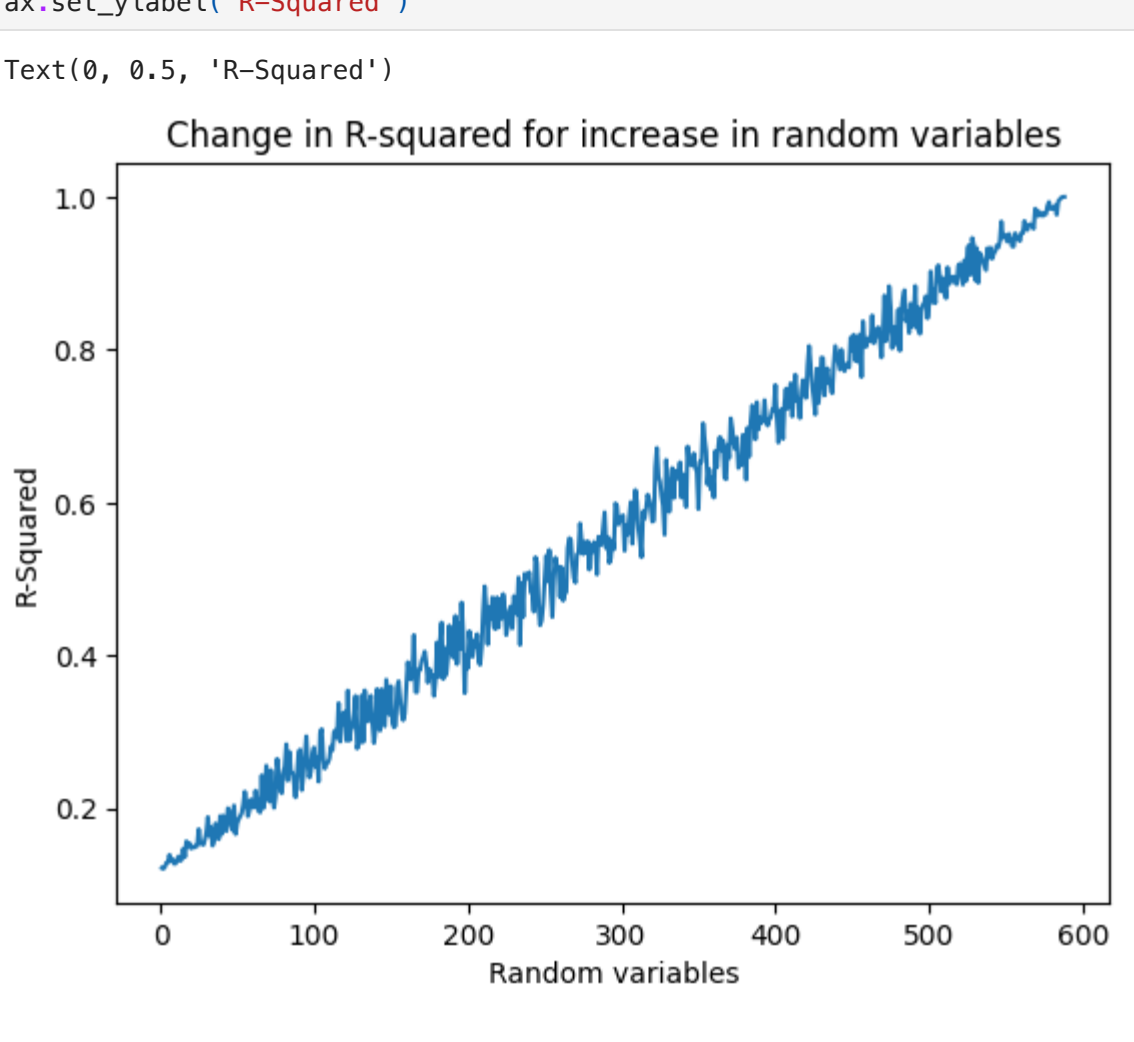
```
# I thought it would be cool to visualize how our R-square changes for each additional Random variable we creat
results = []
for i in range(1,589):
    frame = random_variables(i)
    columns = frame.columns
    seperator = " + "
    string = seperator.join(columns[8:(9+i)])
    mod = sm.ols(f"life_exp ~ cur_smoke + {string}", data=frame)
    res = mod.fit(cov_type='HC2')
    results.append({'Variables': i, 'R-squared': res.rsquared})
```

In [11]:

```
vis_frame = pd.DataFrame(results)
vis_frame.head()
x = vis_frame['Variables']
y = vis_frame['R-squared']

fig, ax = plt.subplots(1, 1)
plt.plot(x, y)
ax.set_title('Change in R-squared for increase in random variables')
ax.set_xlabel('Random variables')
ax.set_ylabel('R-Squared')
```

Out[11]: Text(0, 0.5, 'R-Squared')



Question 5.

Explain (d) using the following simple non-trivial example: with N = 2 observations, what happens if we run a regression with an intercept and one explanatory variable?

In order to simulate this we shall create a random data set with N = 2 observations. Hence, we change our helper function to accept N as an input.

In [7]:

```
def random_variables_df(x, N):
#     N = health.shape[0]
    # Create column names like rand_5, rand_6...
    random_column_names = [f"rand_{i}" for i in range(x)]

    # Create a new dataframe with just random columns
    random_df = pd.DataFrame(
        np.random.random(size=(N, x)),
        columns=random_column_names
    )
    return random_df

frame_5 = random_variables_df(2, 2)
frame_5.head()
```

Out[7]:

	rand_0	rand_1
0	0.477313	0.422674
1	0.463580	0.268477

In [13]:

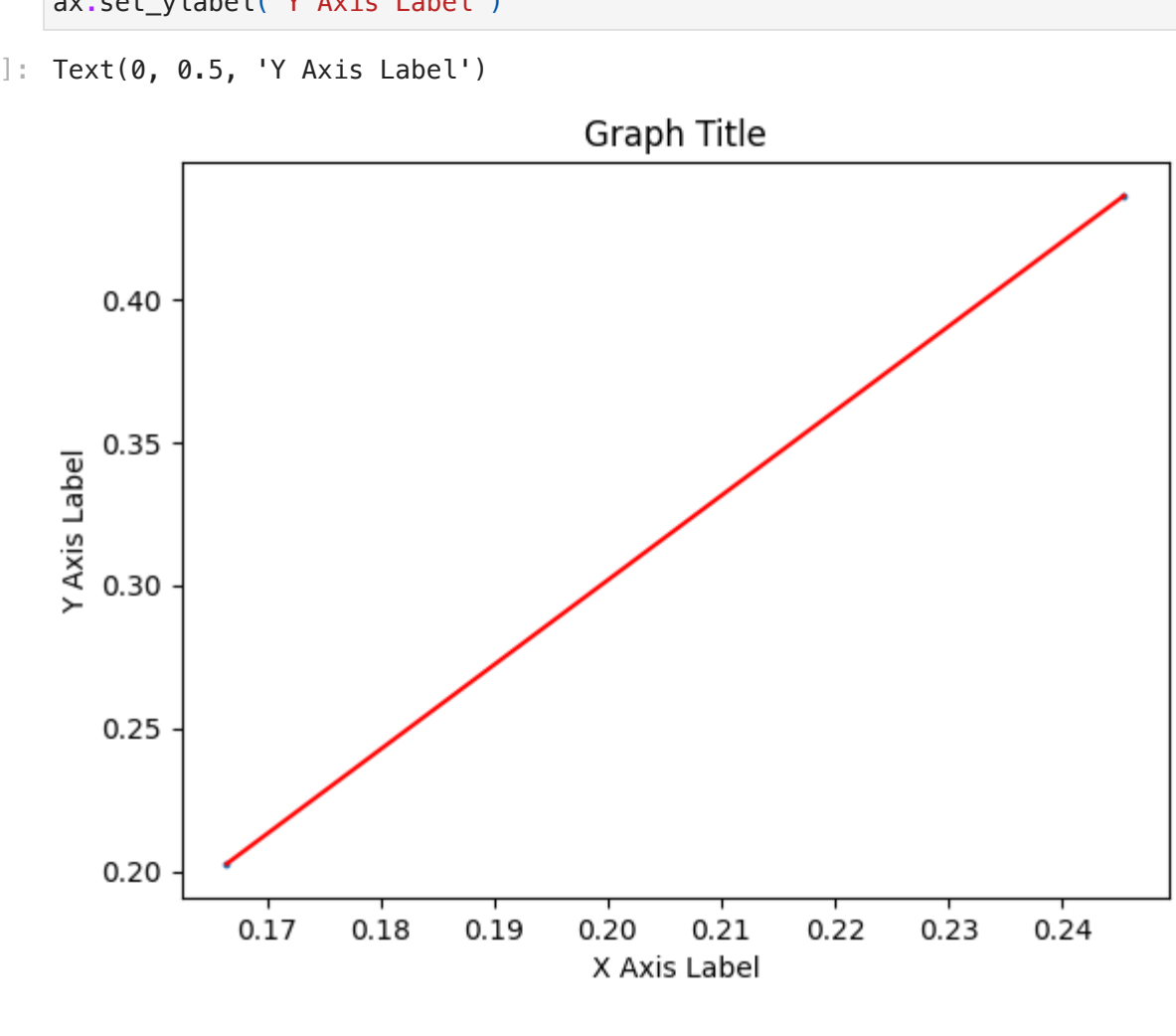
```
# mod = sm.ols('rand_0 ~ rand_1', data=frame_5).fit(cov='HC2')
x1 = frame_5['rand_1']
y1 = frame_5['rand_0']
# Sets up space for graphing
fig, ax = plt.subplots(1, 1)

# Plots scatter plot (size of points is set by s=3)
ax.scatter(x1, y1, s=3)

# Adds a line best fit for the data to the plot
ax.plot(
    *np.polynomial.Polynomial.fit(
        x1, y1, 1
    ).linspace(2),
    "r-"
)

# Label axes
ax.set_title("Graph Title")
ax.set_xlabel("X Axis Label")
ax.set_ylabel("Y Axis Label")
```

Out[13]: Text(0, 0.5, 'Y Axis Label')



If we simply only posess two observations then our residuals will be zero, since our simple linear regression of degree one. A linear function will be going through both points.

Question 6.

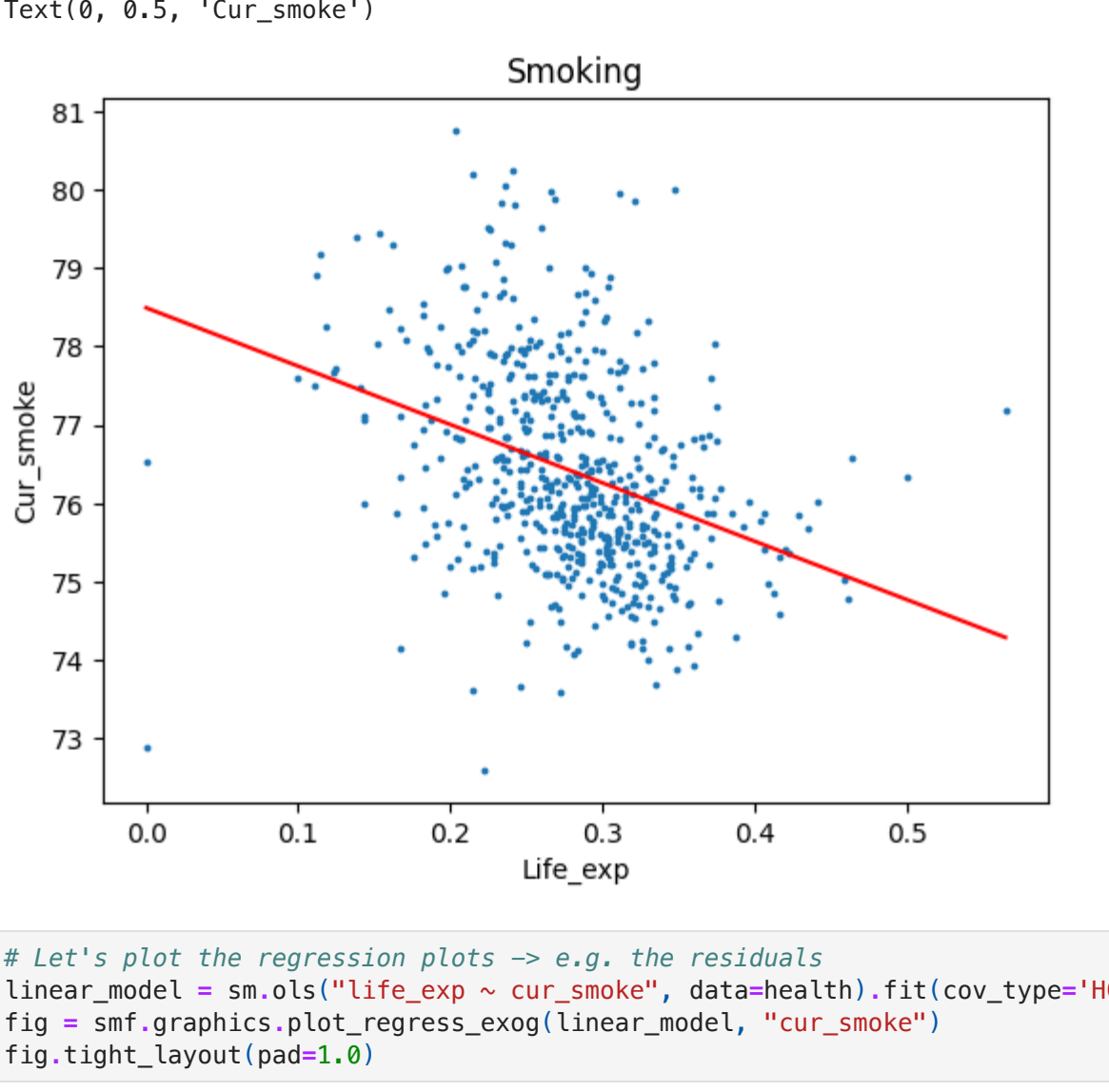
Now generate a total of 589 random variables. Regress life_exp on an intercept, cur_smoke, and the 589 independent random variable that you generated. What happens?

Nothing happens to the R-squared since, it already is at 1.0. Hence, saying that our model already explains all the variance in the dependent variable.

In [27]:

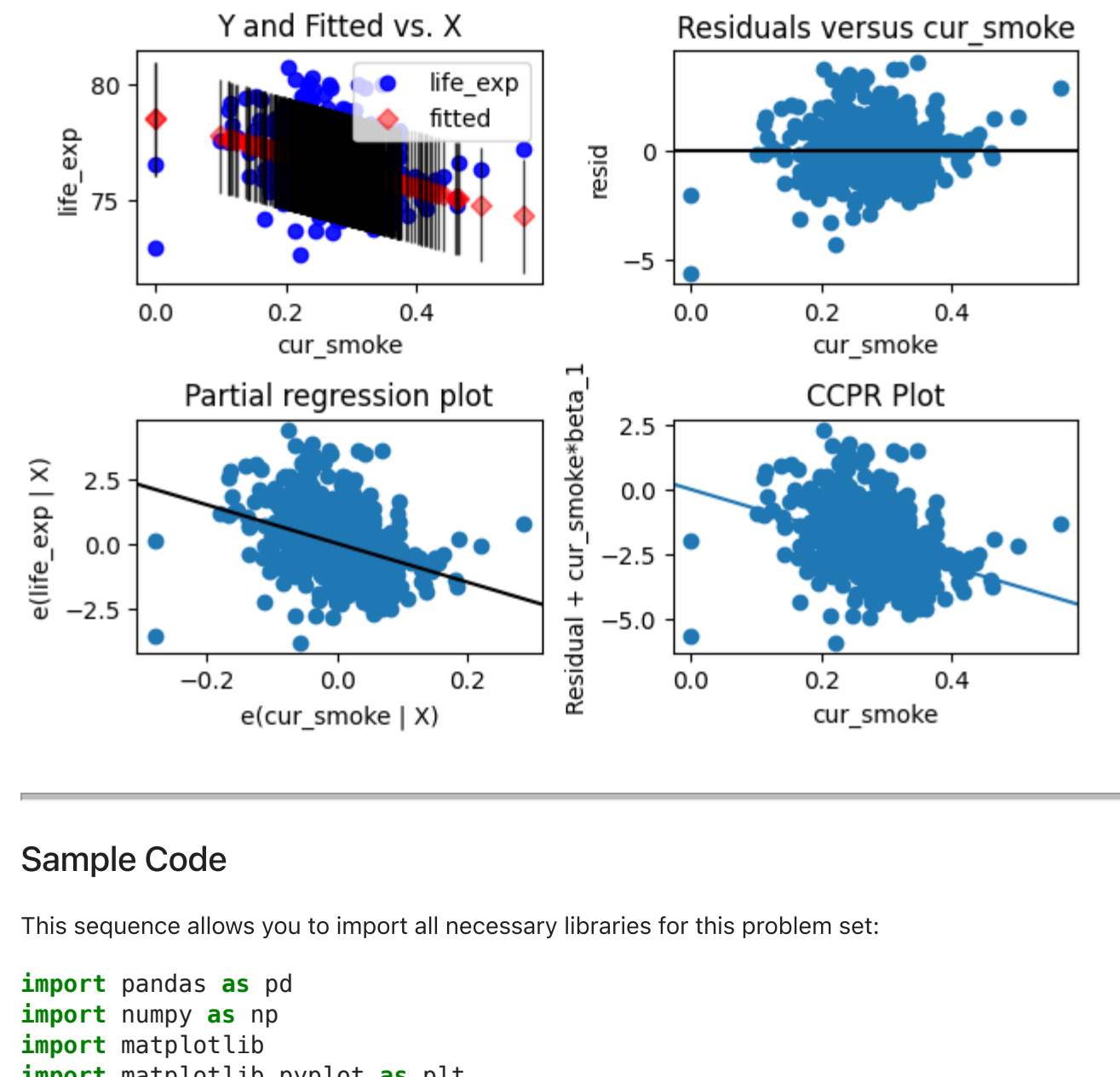
```
# Let's not print the output, cause it's so long
frame = random_variables(589)
columns = frame.columns
seperator = " + "
string = seperator.join(columns[8:])
mod_3 = sm.ols(f"life_exp ~ cur_smoke + {string}", data=frame)
res_3 = mod_3.fit(cov_type='HC2')
# print(res_3.summary(slim=True))
```

Out[17]: Text(0, 0.5, 'Cur_smoke')



In [32]:
```python
# Let's plot the regression plots -> e.g. the residuals
linear_model = sm.ols("life_exp ~ cur_smoke", data=health).fit(cov_type='HC2')
fig = smf.graphics.plot_regress_exog(linear_model, "cur_smoke")
fig.tight_layout(pad=1.0)
```

eval_env: 1



Sample Code

This sequence allows you to import all necessary libraries for this problem set:

```python
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import statsmodels.formula.api as sm
# Uncomment the line below if using Google Colab:
# !pip install stargazer
from stargazer.stargazer import Stargazer
```

Reads in data and displays first 5 rows in dataset. We add an intercept column because Python does not automatically add a constant to regressions like Stata and R:

```python
# Read dataset into a pandas dataframe
health = pd.read_stata("healthinequality.dta")

# Display first 5 rows of data
health.head()
```

Estimates regression of yvar on an intercept and xvar1, with HC2 heteroskedasticity-robust standard errors. `cov_type="HC2"` corresponds to HC2 `cov_type="HC1"` corresponds to HC1

```python
mod = sm.ols(
    "yvar ~ xvar",
    data=health
)
res = mod.fit(cov_type="HC2")

# print results:
res.summary(slim=True)
```

Create x columns of random variables and add them to a new dataframe

```python
# Find number of rows
N = health.shape[0]

# Create column names like rand_5, rand_6...
random_column_names = [f"rand_{i}" for i in range(x)]

# Create a new dataframe with just random columns
random_df = pd.DataFrame(
    np.random.random(size=(N, x)),
    columns=random_column_names
)

# Join old and new dataframes
new_df = pd.concat([health, random_df], axis=1)

# Extra hint: look at the python "join" function to
# create a string out of the list of variable names
```

Create regression table with custom column labels

```python
# Estimate Regressions:
mod1 = sm.ols(
“yvar1 ~ xvar1 + xvar2 + xvar3”,
data=health
)
res1 = mod.fit(cov_type=”HC2”)
mod2 = sm.ols(
“yvar1 ~ xvar1 + xvar2”,
data=health
)
res2 = mod.fit(cov_type=”HC2”)
mod3 = sm.ols(
“yvar2 ~ xvar2 + xvar3”,
data=health
)
res3 = mod.fit(cov_type=”HC2”)

# Create Table
table = Stargazer(models)

# Label columns
# This list of 1s should be the same length as the
# number of columns
table.custom_columns([“yvar1”, “yvar1”, “yvar2”],
seperators=[1, 1, 1])

# Display table
table
```

Calculate residuals from regression of variable $yvar$ on variables $xvar1$ and $xvar2$

```python
# Estimate Regressions:
mod1 = sm.ols(
    "yvar ~ xvar1 + xvar2",
    data=health
)
res1 = mod.fit(cov_type="HC2")

# Find residuals
residuals = res1.resid
```

Draw a scatter plot of variable $x1$ against variable $y1$ and add a line best fit

```python
# Sets up space for graphing
fig, ax = plt.subplots(1, 1)

# Plots scatter plot (size of points is set by s=3)
ax.scatter(x1, y1, s=3)

# Adds a line best fit for the data to the plot
ax.plot(
    *np.polynomial.Polynomial.fit(
        x1, y1, 1
    ).linspace(2),
    "r-"
)

# Label axes
ax.set_title("Graph Title")
ax.set_xlabel("X Axis Label")
ax.set_ylabel("Y Axis Label")
```