In [ ]:
```
import pandas as pd
df = pd.read_stata("A1_kommune.dta")
print(df)
```

```
    nr                kommune        taxrev    taxrate     pop
0   101     Københavns Kommune  44170.335938  23.799999  528208
1   147  Frederiksberg Kommune   6682.439941  23.100000   96718
2   151       Ballerup Kommune   4598.704102  25.500000   47652
3   153        Brøndby Kommune   3121.204834  24.500000   33795
4   155         Dragør Kommune    858.017944  24.799999   13564
..  ...                    ...           ...        ...     ...
93  840         Rebild Kommune   1957.348511  25.100000   28852
94  846  Mariagerfjord Kommune   3186.054932  25.700001   42604
95  849     Jammerbugt Kommune   3241.256592  25.299999   38927
96  851        Aalborg Kommune  16330.091797  25.400000  197426
97  860       Hjørring Kommune   5061.796875  25.400000   66803

[98 rows x 5 columns]
```

In [ ]:
```
df[['taxrev', 'taxrate', 'pop']].describe()
```

Out[ ]:

	taxrev	taxrate	pop
count	98.000000	98.000000	98.000000
mean	4477.341309	25.208162	56475.887755
std	5251.175293	0.908003	62925.301713
min	211.228409	22.799999	1969.000000
25%	2466.702271	24.799999	29997.750000
50%	3317.848633	25.299999	43475.000000
75%	4786.060913	25.700001	59733.000000
max	44170.335938	27.799999	528208.000000

In [ ]:
```
df['kommune'].value_counts(normalize=True)
```

Out[ ]:
```
kommune
Københavns Kommune    0.010204
Syddjurs Kommune      0.010204
Lemvig Kommune        0.010204
Holstebro Kommune     0.010204
Herning Kommune       0.010204
                        ...
Køge Kommune          0.010204
Greve Kommune         0.010204
Bornholms Kommune     0.010204
Gribskov Kommune      0.010204
Hjørring Kommune      0.010204
Name: proportion, Length: 98, dtype: float64
```

In [ ]:
```
import numpy as np
import statsmodels.api as sm

y = np.log(df['taxrev'])
X = df['taxrate']
```

```
X = sm.add_constant(X)

OLS = sm.OLS(y, X).fit()
print(OLS.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  taxrev   R-squared:
0.029
Model:                            OLS   Adj. R-squared:
0.018
Method:                 Least Squares   F-statistic:
2.818
Date:                Thu, 18 Sep 2025   Prob (F-statistic):             0.
0965
Time:                        14:05:09   Log-Likelihood:                -11
1.12
No. Observations:                  98   AIC:                             2
26.2
Df Residuals:                      96   BIC:                             2
31.4
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.
975]
------------------------------------------------------------------------------
const         11.6982      2.143      5.459      0.000       7.444      1
5.952
taxrate       -0.1426      0.085     -1.679      0.096      -0.311
0.026
==============================================================================
Omnibus:                       17.794   Durbin-Watson:
1.863
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                4
5.446
Skew:                          -0.567   Prob(JB):                     1.35
e-10
Kurtosis:                       6.138   Cond. No.
705.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
```

In [ ]:
```
df['logtaxrev'] = np.log(df['taxrev'])

y = df['logtaxrev']
df['logpop'] = np.log(df['pop'])

X = sm.add_constant(df[['taxrate', 'logpop']])

OLS = sm.OLS(y, X).fit()
print(OLS.summary())
```

```
                            OLS Regression Results
==============================================================================
====
Dep. Variable:              logtaxrev   R-squared:
0.980
Model:                            OLS   Adj. R-squared:
0.980
Method:                 Least Squares   F-statistic:                      2
344.
Date:                Thu, 18 Sep 2025   Prob (F-statistic):           1.42
e-81
Time:                        14:05:09   Log-Likelihood:                  7
9.497
No. Observations:                  98   AIC:                             -1
53.0
Df Residuals:                      95   BIC:                             -1
45.2
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.
975]
------------------------------------------------------------------------------
----
const         -2.8022      0.376     -7.461      0.000      -3.548      -
2.057
taxrate        0.0226      0.012      1.816      0.072      -0.002
0.047
logpop         0.9711      0.014     67.471      0.000       0.943
1.000
==============================================================================
====
Omnibus:                       15.730   Durbin-Watson:
2.007
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                2
0.068
Skew:                           0.810   Prob(JB):                     4.39
e-05
Kurtosis:                       4.514   Cond. No.
933.
==============================================================================
====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
```

In [ ]:

```python
import pandas as pd

corr = df[["taxrate", "logpop"]].corr()


cov = df[["taxrate", "logpop"]].cov()

print("Korrelation:")
print(corr, "\n")

print("Kovarians:")
print(cov)
```

```
Korrelation:
          taxrate    logpop
taxrate  1.000000 -0.196649
logpop  -0.196649  1.000000

Kovarians:
          taxrate    logpop
taxrate  0.824469 -0.140289
logpop  -0.140289  0.617290
```

In [ ]:

```
X = sm.add_constant(df['logpop'])
y = df['taxrate']


results = sm.OLS(y, X).fit()
df['res1'] = results.resid

print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
====
Dep. Variable:                taxrate   R-squared:
0.039
Model:                            OLS   Adj. R-squared:
0.029
Method:                 Least Squares   F-statistic:
3.862
Date:                Thu, 18 Sep 2025   Prob (F-statistic):             0.
0523
Time:                        14:05:57   Log-Likelihood:                -12
7.16
No. Observations:                  98   AIC:                             2
58.3
Df Residuals:                      96   BIC:                             2
63.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.
975]
------------------------------------------------------------------------------
----
const         27.6268      1.234     22.386      0.000      25.177      3
0.077
logpop        -0.2273      0.116     -1.965      0.052      -0.457
0.002
==============================================================================
====
Omnibus:                        3.367   Durbin-Watson:
1.570
Prob(Omnibus):                  0.186   Jarque-Bera (JB):
2.699
Skew:                          -0.328   Prob(JB):
0.259
Kurtosis:                       3.480   Cond. No.
147.
==============================================================================
====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
                            OLS Regression Results
==============================================================================
====
Dep. Variable:              logtaxrev   R-squared:
0.001
Model:                            OLS   Adj. R-squared:                 -
0.010
Method:                 Least Squares   F-statistic:                    0.0
6626
Date:                Thu, 18 Sep 2025   Prob (F-statistic):
0.797
Time:                        14:05:57   Log-Likelihood:                -11
2.50
No. Observations:                  98   AIC:                             2
29.0
```

```
Df Residuals:                       96   BIC:                             2
34.2
Df Model:                            1
Covariance Type:            nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.
975]
------------------------------------------------------------------------------
----
const          8.1031      0.078    104.100      0.000       7.949
8.258
res1           0.0226      0.088      0.257      0.797      -0.152
0.197
==============================================================================
====
Omnibus:                       15.581   Durbin-Watson:
1.763
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                4
6.400
Skew:                          -0.383   Prob(JB):                     8.40
e-11
Kurtosis:                       6.283   Cond. No.
1.13
==============================================================================
====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
```

In [ ]:
```
y1 = df['logtaxrev']
X1 = sm.add_constant(df['res1'])

results2 = sm.OLS(y1, X1).fit()
print(results2.summary())
```

```
                            OLS Regression Results
==============================================================================
====
Dep. Variable:              logtaxrev   R-squared:
0.001
Model:                            OLS   Adj. R-squared:                 -
0.010
Method:                 Least Squares   F-statistic:                    0.0
6626
Date:                Thu, 18 Sep 2025   Prob (F-statistic):
0.797
Time:                        14:08:13   Log-Likelihood:                -11
2.50
No. Observations:                  98   AIC:                             2
29.0
Df Residuals:                      96   BIC:                             2
34.2
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.
975]
------------------------------------------------------------------------------
----
const          8.1031      0.078    104.100      0.000       7.949
8.258
res1           0.0226      0.088      0.257      0.797      -0.152
0.197
==============================================================================
====
Omnibus:                       15.581   Durbin-Watson:
1.763
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               4
6.400
Skew:                          -0.383   Prob(JB):                    8.40
e-11
Kurtosis:                       6.283   Cond. No.
1.13
==============================================================================
====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is cor
rectly specified.
```

In [ ]: