Computer Output

Wooldridge Computer Exercise C8.2

```
                OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.672
Model:                            OLS   Adj. R-squared:                  0.661
Method:                 Least Squares   F-statistic:                     19.54
Date:                Fri, 17 Nov 2023   Prob (F-statistic):           1.06e-09
Time:                        13:42:42   Log-Likelihood:                -482.88
No. Observations:                  88   AIC:                             973.8
Df Residuals:                      84   BIC:                             983.7
Df Model:                           3
Covariance Type:                  HC3
==============================================================================
              coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const     -21.7703     41.033     -0.531      0.596    -102.193      58.652
lotsize     0.0021      0.007      0.289      0.772      -0.012       0.016
sqrft       0.1228      0.041      3.014      0.003       0.043       0.203
bdrms      13.8525     11.562      1.198      0.231      -8.808      36.513
==============================================================================
Omnibus:                       20.398   Durbin-Watson:                   2.110
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               32.278
Skew:                           0.961   Prob(JB):                     9.79e-08
Kurtosis:                       5.261   Cond. No.                     6.41e+04
==============================================================================
```

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)
[2] The condition number is large, 6.41e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

(i)We assume that standard errors have characteristics of homoskedasticity, where the variance of the error term is constant.
But heteroskedasiticity-robust standard errors sense the heteroskedasticity in data. Given data suggests thatour
heteroskedasticity-robust errors are larger than the usual standard errors, so this means there's heteroskedasticity in data,
and these larger numbers give us better results ans estimates for the coefficients.

```
                OLS Regression Results
==============================================================================
```

| | | | |
|---|---|---|---|
| Dep. Variable: | price | R-squared: | 0.639 |
| Model: | OLS | Adj. R-squared: | 0.626 |
| Method: | Least Squares | F-statistic: | 42.42 |
| Date: | Fri, 17 Nov 2023 | Prob (F-statistic): | 8.67e-17 |
| Time: | 13:42:42 | Log-Likelihood: | 25.417 |
| No. Observations: | 88 | AIC: | -42.83 |
| Df Residuals: | 84 | BIC: | -32.92 |
| Df Model: | 3 | | |
| Covariance Type: | HC3 | | |

==============================================================================

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | -1.4295 | 0.827 | -1.728 | 0.084 | -3.050 | 0.191 |
| lotsize | 0.1695 | 0.053 | 3.181 | 0.001 | 0.065 | 0.274 |
| sqrft | 0.7170 | 0.122 | 5.866 | 0.000 | 0.477 | 0.957 |
| bdrms | 0.0989 | 0.130 | 0.761 | 0.447 | -0.156 | 0.354 |

==============================================================================

| | | | |
|---|---|---|---|
| Omnibus: | 11.198 | Durbin-Watson: | 2.056 |
| Prob(Omnibus): | 0.004 | Jarque-Bera (JB): | 30.826 |
| Skew: | -0.144 | Prob(JB): | 2.02e-07 |
| Kurtosis: | 5.885 | Cond. No. | 386. |

==============================================================================

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

(ii)Since heteroskedasticity-robust error is larger than the usual standard error, we know there may be heteroskedsticity in the residuals.

(iii)Based on our observation, when heteroskedasticity-robust error is greater than the usual standard error, there's heteroskedasticy in the residuals.
Therefore, the significance of the coefficent estimates may be influenced by the heteroskedasticity, making the data unreliable.

Wooldridge Computer Exercise C8.4

| | district | democA | voteA | expendA | expendB \ |
|---|---|---|---|---|---|
| count | 173.000000 | 173.000000 | 173.000000 | 173.000000 | 173.000000 |
| mean | 8.838150 | 0.554913 | 50.502890 | 310.611023 | 305.088531 |
| std | 8.768823 | 0.498418 | 16.784761 | 280.985382 | 306.278351 |
| min | 1.000000 | 0.000000 | 16.000000 | 0.302000 | 0.930000 |
| 25% | 3.000000 | 0.000000 | 36.000000 | 81.634003 | 60.054001 |
| 50% | 6.000000 | 1.000000 | 50.000000 | 242.781998 | 221.529999 |

| | | | | | |
|---|---|---|---|---|---|
| 75% | 11.000000 | 1.000000 | 65.000000 | 457.410004 | 450.716003 |
| max | 42.000000 | 1.000000 | 84.000000 | 1470.673950 | 1548.192993 |

| | prtystrA | lexpendA | lexpendB | shareA |
|---|---|---|---|---|
| count | 173.000000 | 173.000000 | 173.000000 | 173.000000 |
| mean | 49.757225 | 5.025557 | 4.944369 | 51.076546 |
| std | 9.983650 | 1.601602 | 1.571143 | 33.483574 |
| min | 22.000000 | -1.197328 | -0.072571 | 0.094635 |
| 25% | 44.000000 | 4.402246 | 4.095244 | 18.867996 |
| 50% | 50.000000 | 5.492164 | 5.400558 | 50.849903 |
| 75% | 56.000000 | 6.125580 | 6.110837 | 84.255096 |
| max | 71.000000 | 7.293476 | 7.344844 | 99.495003 |

```
                  OLS Regression Results
==============================================================================
Dep. Variable:                y   R-squared:                 0.000
Model:                      OLS   Adj. R-squared:           -0.024
Method:           Least Squares   F-statistic:           7.929e-15
Date:          Fri, 17 Nov 2023   Prob (F-statistic):         1.00
Time:                  13:42:42   Log-Likelihood:          -593.20
No. Observations:           173   AIC:                       1196.
Df Residuals:               168   BIC:                       1212.
Df Model:                     4
Covariance Type:      nonrobust
==============================================================================
             coef    std err        t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const     -7.614e-14     4.736  -1.61e-14     1.000     -9.350      9.350
prtystrA   8.503e-16     0.071   1.19e-14     1.000     -0.141      0.141
democA     8.438e-15     1.407      6e-15     1.000     -2.777      2.777
lexpendA  -5.517e-15     0.392  -1.41e-14     1.000     -0.774      0.774
lexpendB   6.407e-15     0.397   1.61e-14     1.000     -0.785      0.785
==============================================================================
Omnibus:                  6.304   Durbin-Watson:              1.525
Prob(Omnibus):            0.043   Jarque-Bera (JB):           6.030
Skew:                     0.448   Prob(JB):                  0.0491
Kurtosis:                 3.182   Cond. No.                    429.
==============================================================================
```

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

(i)Since our coefficients are close to 0 and the standard errors are greater than the coefficients, the independent variables

may not have an relationship/association with the dependent variable when we also consider statistical insignificance.
Maybe our adjusted R squared is over fitting while R squared is not fitting well.

(ii) LM Statistic: 9.093356486631897
LM p-value: 0.05880790411089757
F-statistic: 2.330112826740852
F p-value: 0.05805750110700962

Wooldridge Computer Exercise C9.4

```
            year     infmort     afdcprt       popul        pcinc  \
count   102.000000  102.000000   102.000000    102.000000    102.000000
mean   1988.500000    9.708824   222.068627   4825.774510  16287.460784
std       1.507407    2.059756   315.724509   5298.512199   3163.157135
min    1987.000000    6.200000    11.000000    454.000000  10301.000000
25%    1987.000000    8.525000    44.750000   1190.000000  14052.750000
50%    1988.500000    9.500000   125.500000   3293.500000  15736.000000
75%    1990.000000   10.400000   236.750000   5778.000000  18277.000000
max    1990.000000   20.700001  2023.000000  29760.000000  25528.000000

          physic     afdcper        d90      lpcinc     lphysic        DC  \
count  102.000000  102.000000  102.000000  102.000000  102.000000  102.000000
mean   201.500000    4.042603    0.500000    9.679919    5.260582    0.019608
std     74.374986    1.472844    0.502469    0.191395    0.279374    0.139333
min    120.000000    1.041667    0.000000    9.239996    4.787492    0.000000
25%    161.000000    2.975885    0.000000    9.550571    5.081404    0.000000
50%    185.500000    3.863402    0.500000    9.663701    5.223051    0.000000
75%    212.000000    4.708885    1.000000    9.813397    5.356586    0.000000
max    615.000000    8.896211    1.000000   10.147532    6.421622    1.000000

          lpopul
count  102.000000
mean     7.989398
std      1.023285
min      6.118097
25%      7.081692
50%      8.099706
75%      8.661535
max     10.300920
                            OLS Regression Results
==============================================================================
Dep. Variable:                infmort   R-squared:                       0.643
Model:                            OLS   Adj. R-squared:                  0.629
```

```
Method:             Least Squares   F-statistic:               43.72
Date:           Fri, 17 Nov 2023   Prob (F-statistic):         6.28e-21
Time:                   13:42:42   Log-Likelihood:           -165.37
No. Observations:            102   AIC:                       340.7
Df Residuals:                 97   BIC:                    353.9
Df Model:                      4
Covariance Type:          nonrobust
==============================================================================
              coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    36.2125      6.727      5.383      0.000      22.862      49.563
lpcinc      -2.3964      0.887     -2.703      0.008      -4.156      -0.637
lphysic     -1.5548      0.773     -2.010      0.047      -3.090      -0.020
lpopul       0.5755      0.137      4.215      0.000       0.305       0.846
DC          13.9632      1.247     11.201      0.000      11.489      16.437
==============================================================================
Omnibus:                    2.832   Durbin-Watson:               1.055
Prob(Omnibus):                0.243   Jarque-Bera (JB):             2.838
Skew:                       0.387   Prob(JB):                   0.242
Kurtosis:                   2.736   Cond. No.                    745.
==============================================================================
```

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

(i)The coefficient is about 13.96, and it's the estimated difference in 'infomort' between DC and and another location.
The size suggests that there's a huge increase in the expected value of 'infomort' in DC compared to another location.
p value is 0, so it's statistically significant, since 0 < 0.05.

(ii) The dummy variable 'DC' in the first equation has additional coefficient of DC and standard errors of DC. The R squared and adjusted
R squared of the first equation are higher than the equation without 'DC', so the one with DC has a better fit and can better explain the variation in 'infmort'.
The one with dummy variable also has statistical significance with the associated p value.

**HW9/HW9.py**

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.api as sm
from statsmodels.compat import lzip
from statsmodels.stats.diagnostic import het_breuschpagan


# Include similar code as HW1 for the basic steps
file_location = "/Users/amyliang/Eco441K/HW9/hprice1.dta"
f2 = "/Users/amyliang/Eco441K/HW9/VOTE1.DTA"
f3 = "/Users/amyliang/Eco441K/HW9/INFMRT.DTA"
df= pd.read_stata(file_location)
new_df = pd.read_stata(f2)
df2 = pd.read_stata(f3)

pd.set_option('display.max_columns', None)

# Wooldridge Computer Exercise C8.2
print("Wooldridge Computer Exercise C8.2")

# Build the model
model = sm.OLS(df['price'], sm.add_constant(df[['lotsize', 'sqrft', 'bdrms']]))
results = model.fit(cov_type='HC3')  #'HC3' is used for heteroskedasticity-robust standard errors

# Print the summary
print(results.summary())
ans1 = """
(i)We assume that standard errors have characteristics of homoskedasticity, where the variance of the error term is constant.
But heteroskedasiticity-robust standard errors sense the heteroskedasticity in data. Given data suggests thatour
heteroskedasticity-robust errors are larger than the usual standard errors, so this means there's heteroskedasticity in data,
and these larger numbers give us better results ans estimates for the coefficients.
"""
print(ans1)
print()

# Build the model in logs
model_in_log = sm.OLS(np.log(df['price']), sm.add_constant(np.log(df[['lotsize', 'sqrft', 'bdrms']])))

# Fit the model in logs
results_in_log = model_in_log.fit(cov_type='HC3')
```

```python

# Print the summary
print(results_in_log.summary())

ans2 = """
(ii)Since heteroskedasticity-robust error is larger than the usual standard error, we know there may be heteroskedsticity in the residuals.
"""
print(ans2)
print()

ans3 = """
(iii)Based on our observation, when heteroskedasticity-robust error is greater than the usual standard error, there's heteroskedasticy in the residuals.
Therefore, the significance of the coefficent estimates may be influenced by the heteroskedasticity, making the data unreliable.
"""
print(ans3)

# Wooldridge Computer Exercise C8.4
print("Wooldridge Computer Exercise C8.4")

print(new_df.describe())

# Find the initial model
model = sm.OLS(new_df['voteA'], sm.add_constant(new_df[['prtystrA', 'democA', 'lexpendA', 'lexpendB']]))
results = model.fit()

# Obtain OLS residuals
resid = results.resid

# Regress residuals on independent variables
residuals_model = sm.OLS(resid, sm.add_constant(new_df[['prtystrA', 'democA', 'lexpendA', 'lexpendB']]))
residuals_results = residuals_model.fit()

# Print results
print(residuals_results.summary())

ans4 = """
(i)Since our coefficients are close to 0 and the standard errors are greater than the coefficients, the independent variables
may not have an relationship/association with the dependent variable when we also consider statistical insignificance.
Maybe our adjusted R squared is over fitting while R squared is not fitting well.
"""

print(ans4)
print()
# Use result from previous OLS regression
residuals = results.resid
```

```python

# Perform the Breusch-Pagan test
lm, lm_p_value, fvalue, f_p_value = het_breuschpagan(residuals, results.model.exog)

# Print the results
print("(ii)",f'LM Statistic: {lm}\nLM p-value: {lm_p_value}\nF-statistic: {fvalue}\nF p-value: {f_p_value}')
print()

# Wooldridge Computer Exercise C9.4
print("Wooldridge Computer Exercise C9.4")

print(df2.describe())
# Get the model with the dummy variable
model_with_dummy = 'infmort ~ lpcinc + lphysic + lpopul + DC'

# Fit the model
results_with_dummy = sm.OLS.from_formula(model_with_dummy, df2).fit()

# Print results
print(results_with_dummy.summary())

ans5 = """
(i)The coefficient is about 13.96, and it's the estimated difference in 'infomort' between DC and and another location.
The size suggests that there's a huge increase in the expected value of 'infomort' in DC compared to another location.
p value is 0, so it's statistically significant, since 0 < 0.05.
"""

print(ans5)
print()

ans6 = """
(ii) The dummy variable 'DC' in the first equation has additional coefficient of DC and standard errors of DC. The R squared and adjusted
R squared of the first equation are higher than the equation without 'DC', so the one with DC has a better fit and can better explain the variation in 'infmort'.
The one with dummy variable also has statistical significance with the associated p value.
"""

print(ans6)
```