In [1]:

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
```

In [2]:

```python
df=pd.read_csv('ai_mid3.csv', header=0)
```

In [3]:

```python
df.head()
```

Out[3]:

| | CRIM | INDUS | RM | TAX | PTRATIO | LSTAT | MEDV |
|---|---|---|---|---|---|---|---|
| 0 | 0.00632 | 2.31 | 6.575 | 296.0 | 15.3 | 4.98 | 24.0 |
| 1 | 0.02731 | 7.07 | 6.421 | 242.0 | 17.8 | 9.14 | 21.6 |
| 2 | 0.02729 | 7.07 | 7.185 | 242.0 | 17.8 | 4.03 | 34.7 |
| 3 | 0.03237 | 2.18 | 6.998 | 222.0 | 18.7 | 2.94 | 33.4 |
| 4 | 0.06905 | 2.18 | 7.147 | 222.0 | 18.7 | 5.33 | 36.2 |

In [7]:

```python
y = df['MEDV']
```

'TAX', 'PTRATIO', 'LSTAT' を説明変数としたモデル
making model with 'TAX', 'PTRATIO', and 'LSTAT' as explanatory variables

In [4]:

```python
x1 = sm.add_constant(df[['TAX', 'PTRATIO', 'LSTAT']])
```

In [8]:

```python
model1 = sm.OLS(y, x1)
result1=model1.fit()
```

```
print(result1.summary())
```

```
                          OLS Regression Results
==============================================================================
Dep. Variable:                   MEDV   R-squared:                       0.607
Model:                            OLS   Adj. R-squared:                  0.604
Method:                 Least Squares   F-statistic:                     258.1
Date:                Fri, 12 Jun 2020   Prob (F-statistic):           2.70e-101
Time:                        10:14:41   Log-Likelihood:                 -1604.2
No. Observations:                 506   AIC:                             3216.
Df Residuals:                     502   BIC:                             3233.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         54.0748      2.263     23.894      0.000      49.628      58.521
TAX            0.0002      0.002      0.096      0.923      -0.004       0.004
PTRATIO       -1.1496      0.136     -8.461      0.000      -1.417      -0.883
LSTAT         -0.8221      0.044    -18.869      0.000      -0.908      -0.736
==============================================================================
Omnibus:                      149.686   Durbin-Watson:                   0.995
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              399.112
Skew:                           1.457   Prob(JB):                     2.16e-87
Kurtosis:                       6.230   Cond. No.                     3.90e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly s
pecified.
[2] The condition number is large, 3.9e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

## Ans.1

adjusted $R^2$ = **0.604**
coef. of LSTAT = **-0.822**

'CRIM', 'INDUS', 'RM' を説明変数としたモデル作成。
making model with 'CRIM', 'INDUS', 'RM' as explanatory variables

```
x2 = sm.add_constant(df[['CRIM', 'INDUS', 'RM']])
```

```
model2 = sm.OLS(y, x2)
result2=model2.fit()
```

```
print(result2.summary())
```

```
                          OLS Regression Results
==============================================================================
Dep. Variable:                   MEDV   R-squared:                       0.566
Model:                            OLS   Adj. R-squared:                  0.563
Method:                 Least Squares   F-statistic:                     217.8
Date:                Fri, 12 Jun 2020   Prob (F-statistic):           1.82e-90
Time:                        10:29:24   Log-Likelihood:                -1629.3
No. Observations:                 506   AIC:                             3267.
Df Residuals:                     502   BIC:                             3284.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -22.1409      2.867     -7.722      0.000     -27.774     -16.508
CRIM           -0.2006      0.035     -5.811      0.000      -0.268      -0.133
INDUS          -0.2394      0.046     -5.217      0.000      -0.330      -0.149
RM              7.6480      0.420     18.227      0.000       6.824       8.472
==============================================================================
Omnibus:                      226.524   Durbin-Watson:                   0.762
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1657.380
Skew:                           1.797   Prob(JB):                         0.00
Kurtosis:                      11.105   Cond. No.                         163.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly s
pecified.
```

## Ans2.

The most effective variable is **RM**

the increment of the target variable = coef. of RM = **7.648**

## Ans.3

AICは小さいほどよいので、3216 < 3267 よりモデル**1**
Smaller AIC is better. 3216 < 3267 ==> model**1**

In [ ]: