

Multiple Regression

Dataset Background:

This dataset contains information collected by the U.S Census Service concerning housing in Boston Mass. The goal is to predict the median house sale price given different parameters.

Dataset Glimpse:

| crim | indus | rm | dis | tax | lstat | medv |
|---------|-------|-------|--------|-----|-------|------|
| 0.00632 | 2.31 | 6.575 | 4.09 | 296 | 4.98 | 24 |
| 0.02731 | 7.07 | 6.421 | 4.9671 | 242 | 9.14 | 21.6 |
| 0.02729 | 7.07 | 7.185 | 4.9671 | 242 | 4.03 | 34.7 |
| 0.03237 | 2.18 | 6.998 | 6.0622 | 222 | 2.94 | 33.4 |
| 0.06905 | 2.18 | 7.147 | 6.0622 | 222 | 5.33 | 36.2 |
| 0.02985 | 2.18 | 6.43 | 6.0622 | 222 | 5.21 | 28.7 |
| 0.08829 | 7.87 | 6.012 | 5.5605 | 311 | 12.43 | 22.9 |
| 0.14455 | 7.87 | 6.172 | 5.9505 | 311 | 19.15 | 27.1 |
| 0.21124 | 7.87 | 5.631 | 6.0821 | 311 | 29.93 | 16.5 |
| 0.17004 | 7.87 | 6.004 | 6.5921 | 311 | 17.1 | 18.9 |

Total Number of Rows: 39

Total Number of Columns: 7

Column Details:

- crim - per capita crime rate by town.
- indus – proportion of non-retail business acres per town.
- rm – average number of rooms per dwelling.
- dis – weighted distances to five Boston employment centers.
- tax – full-value property-tax rate per \$10,000.
- lstat – % lower status of the population.
- medv – Median value of owner-occupied homes in \$1000's.

Main Dependent Variable: medv.

Using SPSS Software EViews, we have analysed the data:

Descriptive Statistics:

| | CRIM | INDUS | RM | DIS | TAX | LSTAT | MEDV |
|--------------|----------|-----------|----------|----------|-----------|----------|----------|
| Mean | 0.586608 | 7.205128 | 6.076949 | 4.653028 | 294.6923 | 13.74462 | 19.79231 |
| Median | 0.637960 | 8.140000 | 5.966000 | 4.453400 | 307.0000 | 13.04000 | 18.90000 |
| Maximum | 1.612820 | 8.140000 | 7.185000 | 6.592100 | 311.0000 | 29.93000 | 36.20000 |
| Minimum | 0.006320 | 2.180000 | 5.456000 | 3.360300 | 222.0000 | 2.940000 | 12.70000 |
| Std. Dev. | 0.484007 | 1.837318 | 0.407130 | 0.911810 | 26.94356 | 6.249469 | 5.875166 |
| Skewness | 0.286564 | -2.087610 | 1.215584 | 0.735126 | -1.893237 | 0.458697 | 1.208626 |
| Kurtosis | 1.790603 | 5.960611 | 4.105661 | 2.277472 | 5.146716 | 2.976109 | 4.099377 |
| Jarque-Bera | 2.910568 | 42.57123 | 11.59123 | 4.360996 | 30.78689 | 1.368547 | 11.45907 |
| Probability | 0.233334 | 0.000000 | 0.003041 | 0.112985 | 0.000000 | 0.504457 | 0.003249 |
| Sum | 22.87771 | 281.0000 | 237.0010 | 181.4681 | 11493.00 | 536.0400 | 771.9000 |
| Sum Sq. Dev. | 8.902003 | 128.2780 | 6.298690 | 31.59314 | 27586.31 | 1484.123 | 1311.668 |
| Observations | 39 | 39 | 39 | 39 | 39 | 39 | 39 |

Inferences:

- The variable crim is slightly right skewed, ranging between 0.006 to 1.61%.
- The variable indus is left skewed, ranging between 2.18 to 8.14 acres.
- The variable rm is right skewed, ranging between 5.45 to 7.18 rooms.
- The variable dis is slightly right skewed, ranging between 3.36 to 6.59.
- The variable tax is left skewed, ranging between 222 to 311 per \$10,000.
- The variable lstat is slightly right skewed, ranging between 2.94 to 29.93%.
- The variable medv is right skewed, ranging between 12.7 to 36 thousand dollars.
- There is no missing data.

Correlation Analysis:

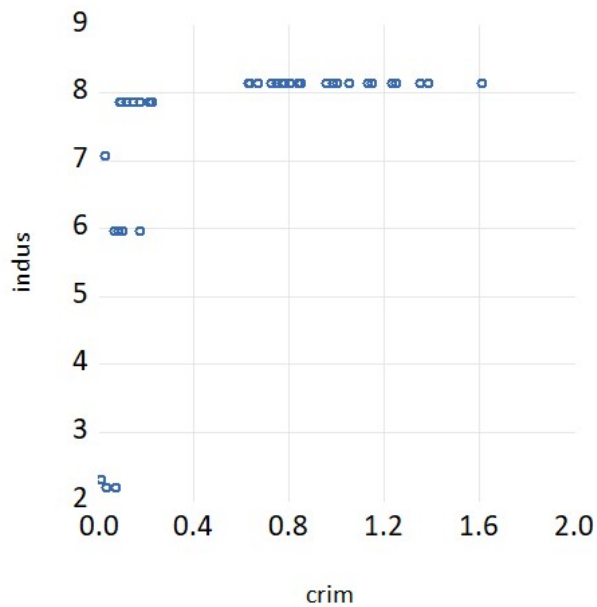
| | CRIM | INDUS | RM | DIS | TAX | LSTAT | MEDV |
|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| CRIM | 1.000000 | 0.571663 | -0.383122 | -0.549310 | 0.523306 | 0.482750 | -0.674332 |
| INDUS | 0.571663 | 1.000000 | -0.589047 | -0.255519 | 0.803176 | 0.586519 | -0.699426 |
| RM | -0.383122 | -0.589047 | 1.000000 | 0.398990 | -0.673633 | -0.531139 | 0.732827 |
| DIS | -0.549310 | -0.255519 | 0.398990 | 1.000000 | -0.257973 | -0.032610 | 0.437578 |
| TAX | 0.523306 | 0.803176 | -0.673633 | -0.257973 | 1.000000 | 0.614706 | -0.761970 |
| LSTAT | 0.482750 | 0.586519 | -0.531139 | -0.032610 | 0.614706 | 1.000000 | -0.727032 |
| MEDV | -0.674332 | -0.699426 | 0.732827 | 0.437578 | -0.761970 | -0.727032 | 1.000000 |

Inferences:

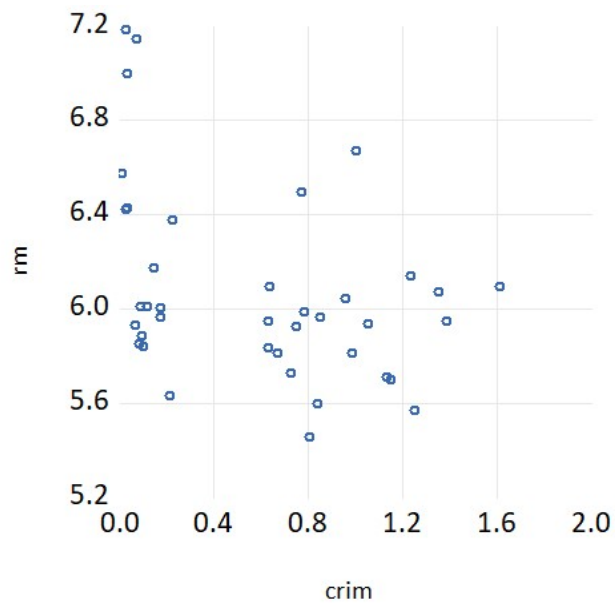
- The variables crim and indus have only the possibility of positive linear correlation, having correlation coefficient 0.57.
- The variables crim and rm have only the possibility of negative linear correlation, having correlation coefficient -0.38.
- The variables crim and dis have only the possibility of negative linear correlation, having correlation coefficient -0.54.
- The variables crim and tax have only the possibility of positive linear correlation, having correlation coefficient 0.52.
- The variables crim and lstat have only the possibility of positive linear correlation, having correlation coefficient 0.48.
- The variables crim and medv have moderate degree of negative linear correlation, having correlation coefficient -0.67.
- The variables indus and rm have only the possibility of negative linear correlation, having correlation coefficient -0.58.
- The variables indus and dis have possibly no correlation, having correlation coefficient -0.25.
- The variables indus and tax have sufficient high degree of positive linear correlation, having correlation coefficient 0.80.
- The variables indus and lstat have only the possibility of positive linear correlation, having correlation coefficient 0.58.
- The variables indus and medv have moderate negative linear correlation, having correlation coefficient -0.69.
- The variables rm and dis have only the possibility of positive linear correlation, having correlation coefficient 0.39.
- The variables rm and tax have moderate degree of negative correlation, having correlation coefficient -0.67.
- The variables rm and lstat have only the possibility of negative linear correlation, having correlation coefficient -0.53.

- The variables rm and medv have moderate positive linear correlation, having correlation coefficient 0.73.
- The variables dis and tax have possibly no linear correlation, having correlation coefficient -0.25.
- The variables dis and lstat have possibly no linear correlation, having correlation coefficient -0.03.
- The variables dis and medv have only the possibility of positive linear correlation, having correlation coefficient 0.43.
- The variables tax and lstat have moderate degree of positive linear correlation, having correlation coefficient 0.61.
- The variables tax and medv have sufficient high degree of negative linear correlation, having correlation coefficient -0.76.
- The variables lstat and medv have moderate degree of negative linear correlation, having correlation coefficient -0.72.

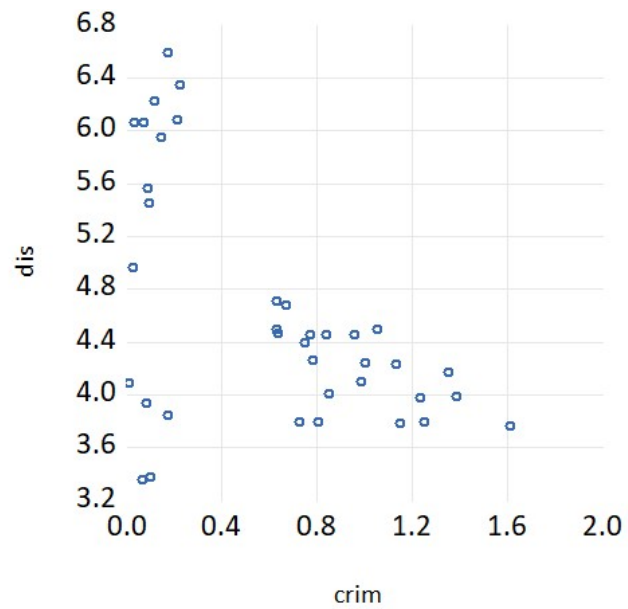
Scatter Plots:

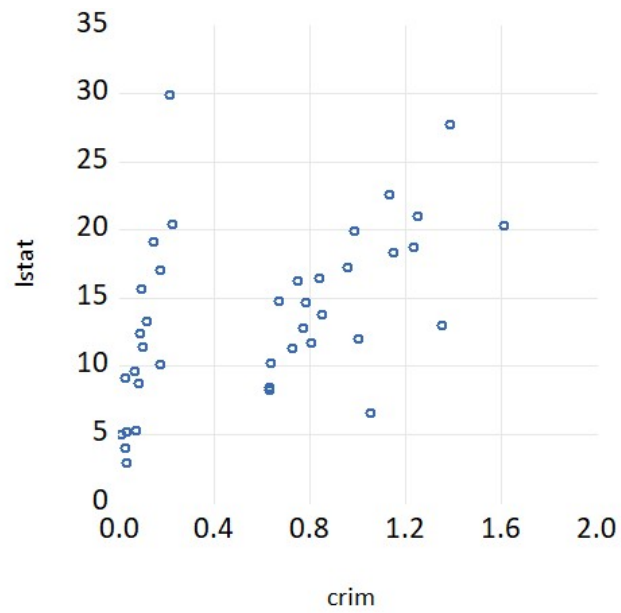


Inference: the variables crim and indus have only the possibility of positive linear correlation.

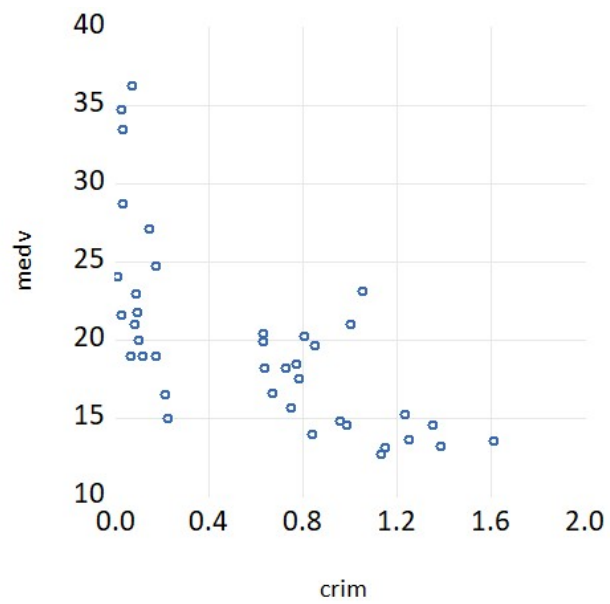


Inference: the variables crim and rm have only the possibility of negative linear correlation.

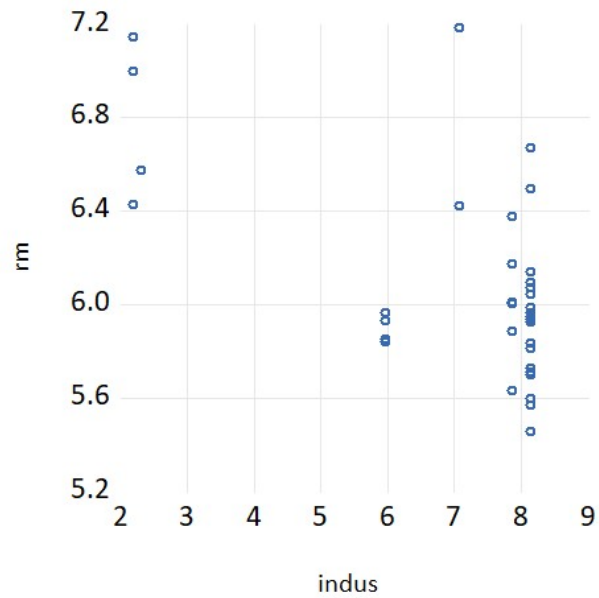




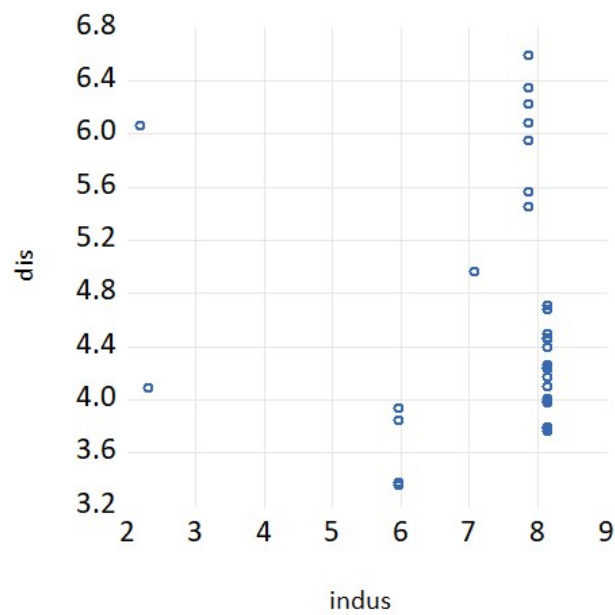
Inference: the variables crim and lstat have only the possibility of positive linear correlation.



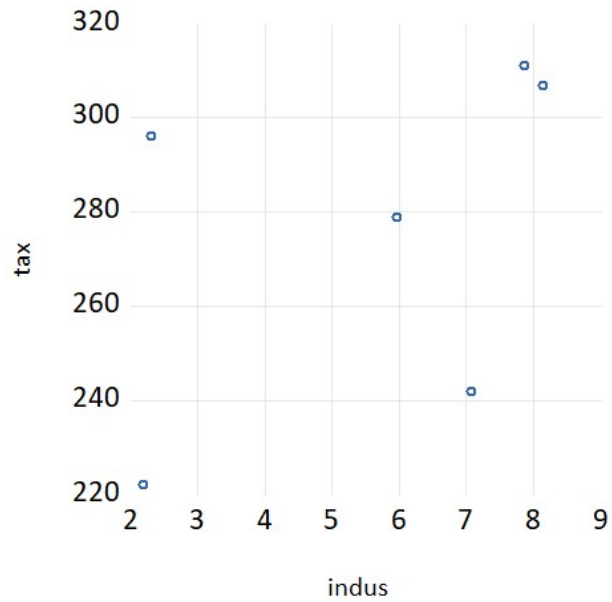
Inference: the variables crim and medv and moderate degree of negative linear correlation.



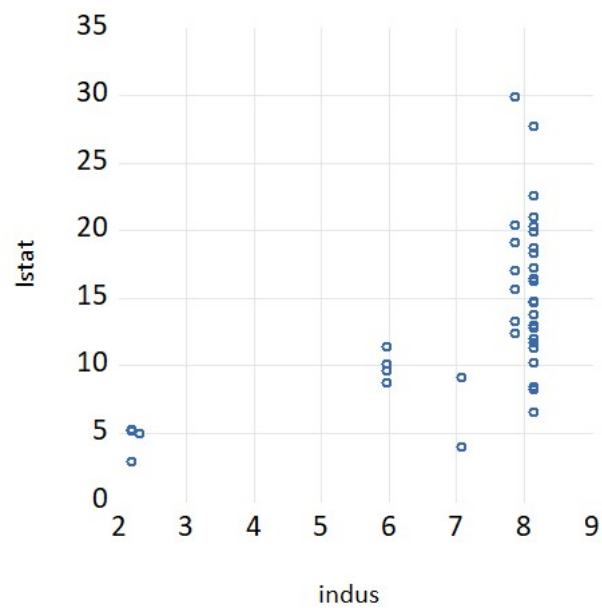
Inference: the variables indus and rm have only the possibility of negative linear correlation.



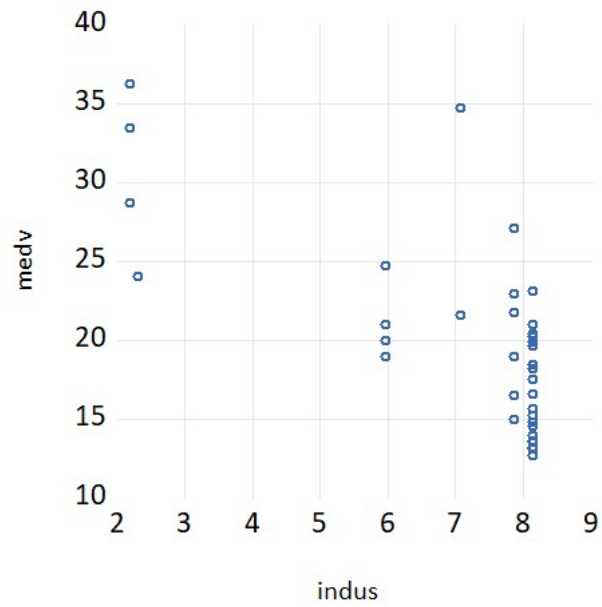
Inference: the variables indus and dis have possibly no correlation.



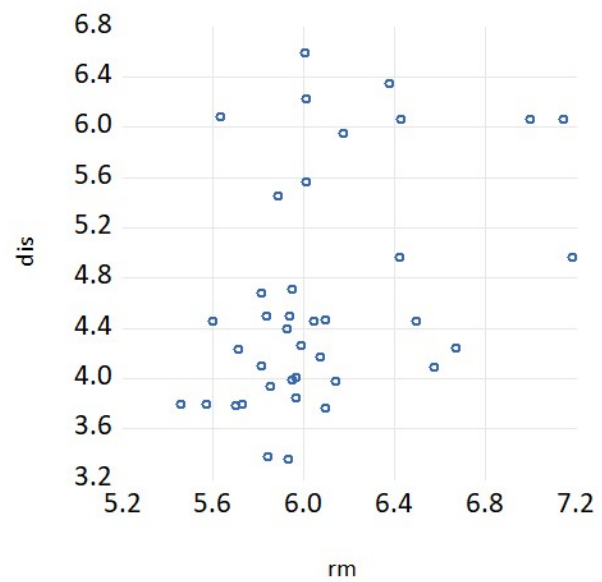
Inference: the variables indus and tax have sufficient high degree of positive linear correlation.



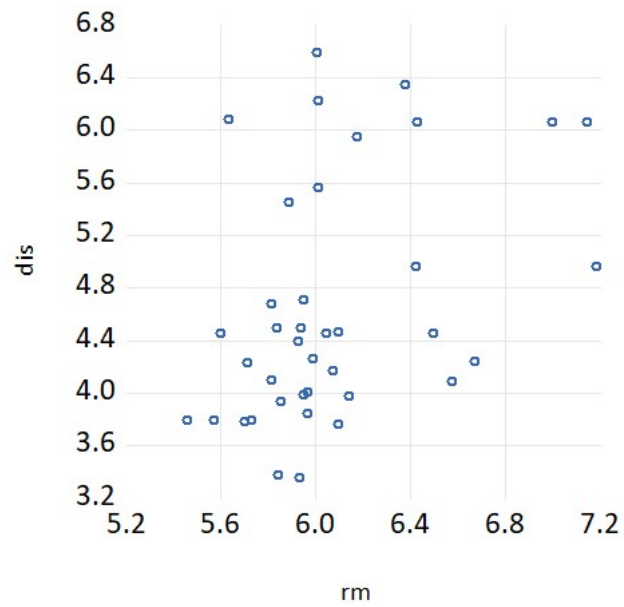
Inference: the variables indus and lstat have only the possibility of positive linear correlation.



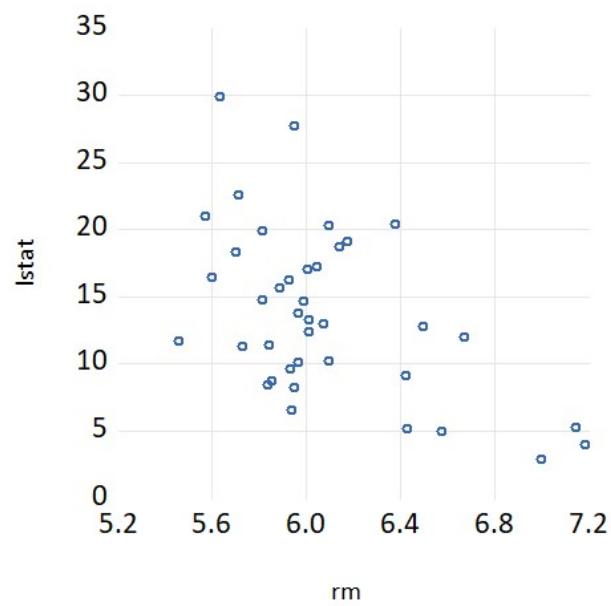
Inference: the variables indus and medv have moderate negative linear correlation.



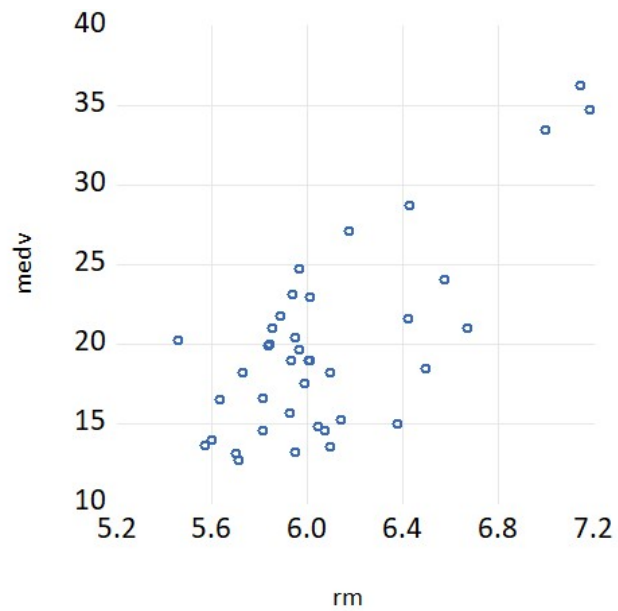
Inference: the variables rm and dis have only the possibility of positive linear correlation.



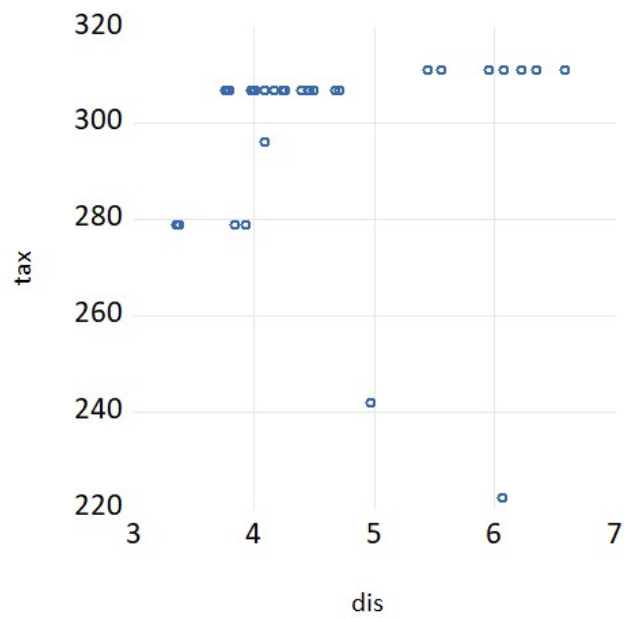
Inference: the variables rm and tax have moderate degree of negative correlation.



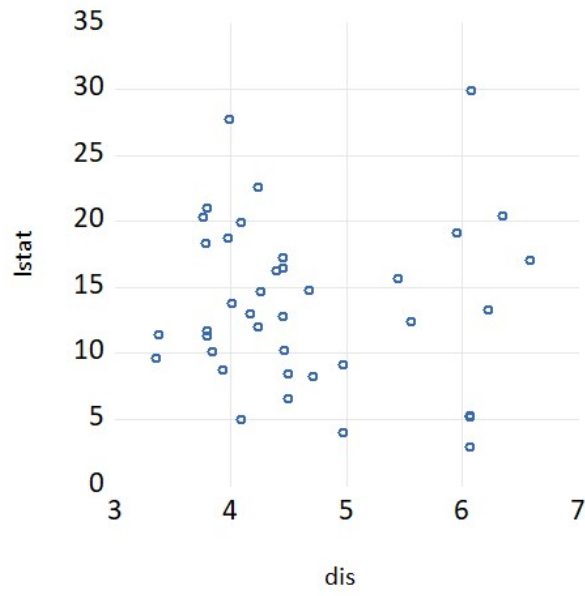
Inference: the variables rm and lstat have only the possibility of negative linear correlation.



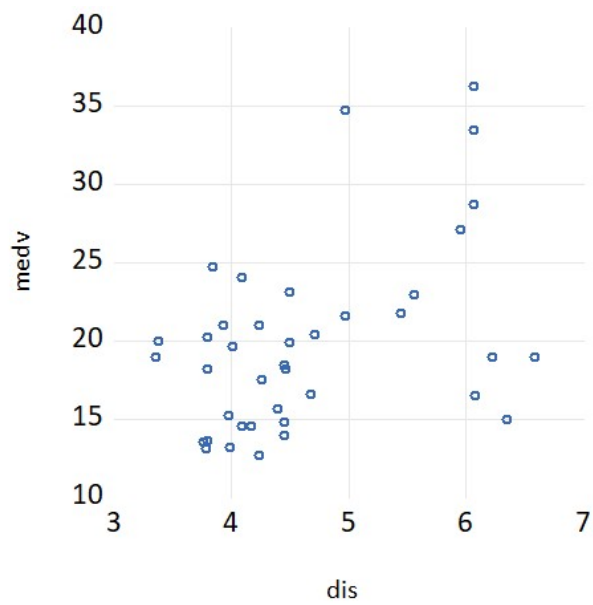
Inference: the variables rm and medv have moderate positive linear correlation.



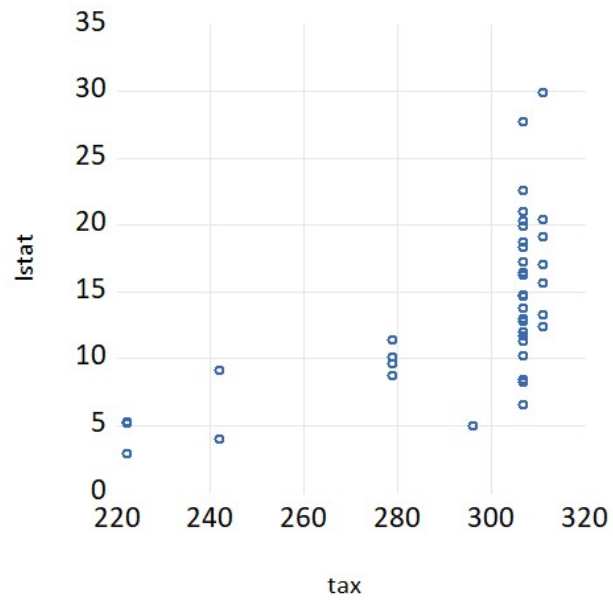
Inference: the variables dis and tax have possibly no linear correlation.



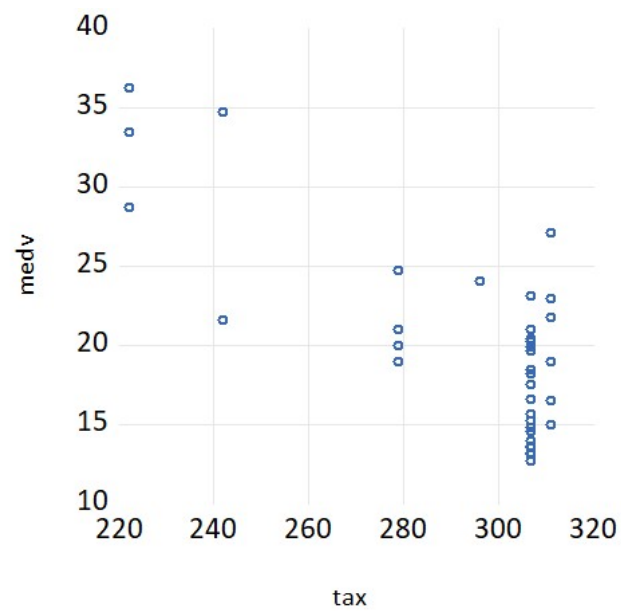
Inference: the variables dis and lstat have possibly no linear correlation.



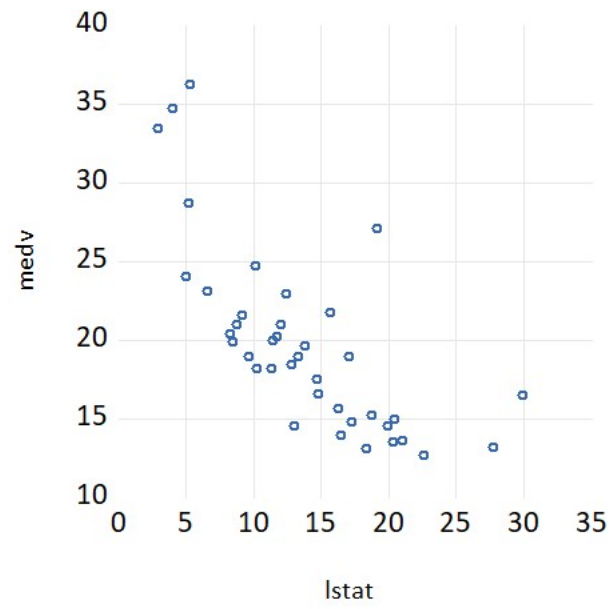
Inference: the variables dis and medv have only the possibility of positive linear correlation.



Inference: the variables tax and lstat have moderate degree of positive linear correlation.



Inference: the variables tax and medv have sufficient high degree of negative linear correlation.



Inference: the variables lstat and medv have moderate degree of negative linear correlation.

Regression Model:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 13.64894 | 14.55694 | 0.937624 | 0.3555 |
| CRIM | -2.445659 | 1.437183 | -1.701704 | 0.0985 |
| INDUS | -0.048574 | 0.429279 | -0.113153 | 0.9106 |
| RM | 3.636100 | 1.675390 | 2.170301 | 0.0375 |
| DIS | 0.990672 | 0.688540 | 1.438801 | 0.1599 |
| TAX | -0.048592 | 0.031051 | -1.564933 | 0.1274 |
| LSTAT | -0.324365 | 0.105116 | -3.085764 | 0.0042 |
| R-squared | 0.819064 | Mean dependent var | 19.79231 | |
| Adjusted R-squared | 0.785139 | S.D. dependent var | 5.875166 | |
| S.E. of regression | 2.723324 | Akaike info criterion | 5.002732 | |
| Sum squared resid | 237.3278 | Schwarz criterion | 5.301320 | |
| Log likelihood | -90.55327 | Hannan-Quinn criter. | 5.109863 | |
| F-statistic | 24.14304 | Durbin-Watson stat | 1.672591 | |
| Prob(F-statistic) | 0.000000 | | | |

Estimate Equation:

$$\text{medv} = 13.64 + (-2.44)(\text{crim}) + (-0.05)(\text{indus}) + (3.64)(\text{rm}) + (0.99)(\text{dis}) + (-0.05)(\text{tax}) + (-0.32)(\text{lstat})$$

Inference:

- The model has an R^2 value of 0.81. Since the value is above 0.75, the model has high explanatory power.
- The adjusted R^2 value of the model is 0.78. Since the value is above 0.75, the model has high explanatory power.
- The variables C, crim, indus, dis, and tax are statistically insignificant with p-values 0.36, 0.1, 0.91, 0.16, and 0.13.
- The variables rm and lstat are statistically significant with p-values 0.038 and 0.0042.