**CST8390 - Lab 7 Answer Document**

**Regression**

**Onur Onel 041074824**

**1.)**

a. Number of instances: **513**

b. Number of attributes: **14**

| ***Attribute*** | ***Type*** | ***Description*** |
| --- | --- | --- |
| CRIM | **Numeric** | Per capita crime rate by town |
| ZN | **Numeric** | Proportion of residential land zoned for lots over 25,000 sq.ft. |
| INDUS | **Numeric** | Proportion of non-retail business acres per town |
| CHAS | **Nominal** | Charles River dummy variable (1 if tract bounds river; 0 otherwise) |
| NOX | **Numeric** | Nitric oxides concentration (parts per 10 million) |
| RM | **Numeric** | Average number of rooms per dwelling |
| AGE | **Numeric** | Proportion of owner-occupied units built prior to 1940 |
| DIS | **Numeric** | Weighted distances to five Boston employment centres |
| RAD | **Numeric** | Index of accessibility to radial highways |
| TAX | **Numeric** | Full-value property-tax rate per $10,000 |
| PTRATIO | **Numeric** | Pupil-teacher ratio by town |
| B | **Numeric** | 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town |
| LSTAT | **Numeric** | % lower status of the population |
| Class | **Numeric** | Median value of owner-occupied homes in $1000's |

**2.)**

The median, being the central value in an ordered list, is determined by sorting the columns in ascending order through the 'Edit' tab. Having done so, I located the middle element at position (513 + 1) / 2, which is the 257th position.

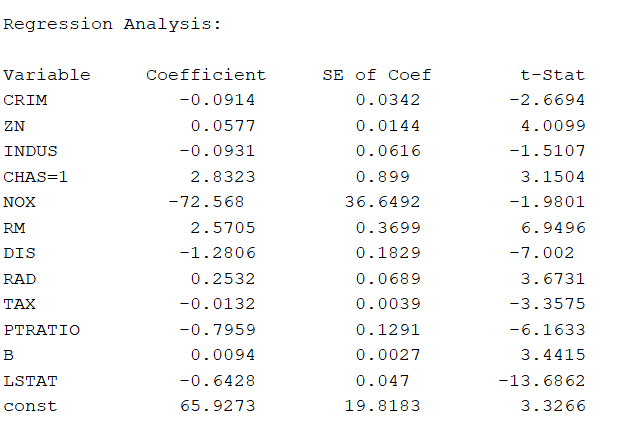
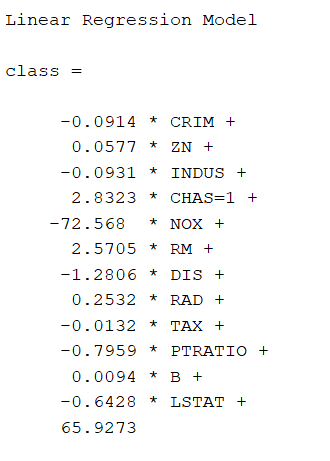
Median House Value (class) x $1000: **21.1** (257th)

Median number of rooms per dwelling:  **8.556** (257th)

Median per capita crime rate: **0.37916** (257th)

**3.)**

**a.** What is the linear regression model for this set?



Description of regression analysis:

1. CRIM (-0.0914): For each unit increase in per capita crime rate, median house value decreases by $91.4, all else being equal.
2. ZN (0.0577): For each unit increase in proportion of residential land zoned for lots over 25,000 sq.ft., median house value increases by $57.7, all else being equal.
3. INDUS (-0.0931): For each unit increase in proportion of non-retail business acres per town, median house value decreases by $93.1, all else being equal.
4. CHAS=1 (2.8323): If the tract bounds Charles River (= 1), the median house value increases by $2832.3, all else being equal.
5. NOX (-72.568): For each unit increase in nitric oxides concentration, median house value decreases by $72,568, all else being equal.
6. RM (2.5705): For each unit increase in average number of rooms per dwelling, median house value increases by $2570.5, all else being equal.
7. DIS (-1.2806): For each unit increase in weighted distances to five Boston employment centres, median house value decreases by $1280.6, all else being equal.
8. RAD (0.2532): For each unit increase in index of accessibility to radial highways, median house value increases by $253.2, all else being equal.
9. TAX (-0.0132): For each unit increase in full-value property-tax rate per $10,000, median house value decreases by $13.2, all else being equal.
10. PTRATIO (-0.7959): For each unit increase in pupil-teacher ratio by town, median house value decreases by $795.9, all else being equal.
11. B (0.0094): For each unit increase in 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town, median house value increases by $9.4, all else being equal.
12. LSTAT (-0.6428): For each unit increase in % lower status of the population, median house value decreases by $642.8, all else being equal.
13. Constant (65.927): This is the baseline median value of owner-occupied homes when all predictors are zero. It is expressed in $1000's, so the baseline value is $65,927.

**b.** The two highest factors that have a positive influence on the housing price are:

CHAS=1 (2.8323): This indicates that if the tract bounds the Charles River (= 1), the median house value increases by $2832.3, all else being equal.

RM (2.5705): This means that for each unit increase in the average number of rooms per dwelling, the median house value increases by $2570.5, all else being equal.

**c.** The two highest factors that have a negative influence on the housing price are:

NOX (-72.568): This indicates that for each unit increase in nitric oxide concentration, the median house value decreases by $72,568, all else being equal.

DIS (-1.2806): This means that for each unit increase in weighted distances to five Boston employment centers, the median house value decreases by $1280.6, all else being equal.

References:

[1] “The Boston Housing Dataset,” Boston Dataset, https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html (accessed Jul. 8, 2023).