Lab 9\_ NWACHUKWU OGOCHUKWU\_GLORIA

6000LEVEL

**EXPLAINED**

**Variables**: The Variables ,"rm", "lstat" and "nox" are the variables that the decision tree found most useful in predicting the target variable (medv).

**residual mean deviance (14.44)-** This is a measure of how well the model fits the data. A lower value indicates that the predictions from the model are closer to the actual values, suggesting a better fit, so it’s a better fit.

**Distribution of residuals** - This part shows the distribution of the differences between the predicted values and the actual values (residuals):

* **Minimum:** -14.92

This is the smallest difference between a predicted value and its corresponding actual value. In this case, the model's prediction could be off by as much as -14.92 units.

* **1st Quartile:** -2.02

The first quartile represents the point below which 25% of the differences between predicted and actual values fall. So, 25% of the model's predictions have differences up to -2.02 units.

* **Median:** -0.01852

The median is the middle value in the dataset. In this context, it's the middle difference between predicted and actual values. It gives you an idea of the typical prediction error. In this case, the median error is -0.01852 units.

* **Mean:** 0.00000

The mean (average) difference between predicted and actual values is very close to zero. Ideally, you want the mean to be as close to zero as possible, as this indicates that, on average, the model is making accurate predictions. In this case, the mean is essentially zero, which is a good sign.

* **3rd Quartile:** 2.18

Similar to the 1st quartile, the 3rd quartile represents the point below which 75% of the differences between predicted and actual values fall. So, 75% of the model's predictions have differences up to 2.18 units.

* **Maximum:** 25.83

This is the largest difference between a predicted value and its corresponding actual value. It represents the maximum potential error of the model, which, in this case, is 25.83 units.

In summary, these statistics provide a summary of how much the model's predictions deviate from the actual values. The mean being close to zero is a positive sign, indicating that, on average, the model is making accurate predictions..

**Lab 9- Plots**

**Regression tree**  
A diagram of a number

Description automatically generated

**To see the tree performance with pruning** A graph with numbers and lines

Description automatically generated

**Pruned**A diagram of a tree

Description automatically generated

**plot(yhat,boston.test)**

A graph with a line and a line

Description automatically generated with medium confidence

**scatter plot comparing the predicted values (yhat.bag) with the actual values in the test set (boston.test)**

A graph with black dots

Description automatically generated

**Used varImpPlot(rf.boston)to see a visual plot of how important each variable is in the prediction**

A graph of a number of individuals

Description automatically generated with medium confidence