**NAME: Sowmy Patlolla**

ASSIGNMENT-2

STAT LEARNING: LINEAR REGRESSION LAB

**SIMPLE LINEAR REGRESSION**

* **TYPICAL USE:**

Simple linear regression in simple words, it is used to determine the relationship between the two quantitative variables and to which degree the correlation between the two quantitative variables can also be identified using simple linear regression.

# **OBSERVATIONS ON THE OUTPUT:**

Based on the given data set which is the boston’s data we can observe that there are certain non-linear characteristics in the data of lstat and medy, we can also see that the data provides the standard errors for the model’s coeffcients and R2 and F- statistic and p-values. From the given data the lm.fit which is present only gives basic information about the model’s output but in order to know more information and conclusion of the data we study the lm.fit data’s summary.

# **MOST SIGNIFICANT FINDING OF THE EXERCISE:**

The most significant finding of the exercise is the lm() function can be used to fit a linear regression model. And another function known as abline () can also be used to add the horizontal and vertical lines. Inorder to increase the width a certain command which is lwd=n is used, when it is used the width will be increased by the factor of n. and there is also an additional option which is pch option which is used for creating several charting symbols.

A screenshot of a computer

Description automatically generated

# **MULTIPLE LINEAR REGRESSION**

* **TYPICAL USE:**

Let’s understand the typical use of multiple linear regression. Suppose if there are two independent variables and one dependent variable in order to establish the relation between these two the multiple linear regression is used. Now we have established the relation between them. Another thing we can do with the help of multiple linear regression is to determine to which degree the dependent variable is connected to more than one independent variable.

# **OBSERVATIONS ON THE OUTPUT:**

So, the observations are it provides to understand that how each coefficient is getting its estimate, standard error and except the age all of them are significant, because the other predictors are interlinked with age hence that is the reason age is not significant and if the predictors are always present then there is no longer the use of age.

# **MOST SIGNIFICANT FINDING OF THE EXERCISE:**

The most significant finding of the exercise is that using in order to fit a multiple linear regression the function called lm() with least squares are used. As there are 12

variables in the Boston data and in order to perform a regression by utilizing the predictors as it is difficult to write all those variables in the data set, so for this task we can use a shorthand command which is lm(medv ~ ., data = boston) and in the same way if would like to perform regression using any specific variables for example age then the same command can used with some changes I,e lm(medv~.,-age).

A screenshot of a computer

Description automatically generated

**QUALITATIVE PREDICTORS**

* **TYPICAL USE:**

So, the typical use of the qualitative predictors is, in the regression analysis, the dependent variables which are present have a lot of impact on various factors such as qualitative, quantitative etc. the linear regression model will use a variable which is dummy and binary in nature that is either zero or one in order to provide qualitative data. The variables which are categorical or qualitative should be included.

# **OBSERVATIONS ON THE OUTPUT:**

Advertising seems to have a substantial impact on income for interaction variables income and contrasts () function is a function which puts two variables in which one is good, and one is dummy.

# **MOST SIGNIFICANT FINDING OF THE EXERCISE:**

The qualitative variables like shelveloc are present in car seats data, a shelveloc is nothing but how well each location’s shelving location. The predictor shelveloc has three possible values which are good bad and medium, The R language creates the dummy variables for the qualitative variables automatically, such as shelveloc. The contrasts () function returns the dummy variables which are used by the R language.

A screenshot of a computer

Description automatically generated

# **WRITING FUNCTIONS:**

# A function in R is an object containing multiple interrelated statements that are run together in a predefined order every time the function is called. Functions in R can be built-in or created by the user (user-defined). The main purpose of creating a user-defined function is to optimize our program, avoid the repetition of the same block of code used for a specific task that is frequently performed in a particular project, prevent us from inevitable and hard-to-debug errors related to copy-paste operations, and make the code more readable. A good practice is creating a function whenever we're supposed to run a certain set of commands more than twice.

A screenshot of a computer

Description automatically generated