*Automobile Insurance Risk Factor Analysis*

Kevin Kuo

Computer Science Department

Towson University

Towson, Maryland

kkuo1@students.towson.edu

*Abstract*— This study analyzes the relationship between a vehicle’s acceleration potential and its perceived insurance risk as well as normalized losses. We have determined there is some positive correlation between sports cars and higher insurance risk and loss.

Keywords— Sports car, car insurance cost, acceleration

# Introduction

It has generally been accepted the sports cars cost more to insure. The general assumption the average consumer or driver has made is that sports car drivers and their vehicles are more likely to experience accidents and therefore carry a higher insurance risk. For the purposes of this study we define a sports car by its acceleration potential. Its acceleration potential is determined primarily by its weight to horsepower ratio. The lower the ratio, the faster the vehicle can accelerate.

# Exploratory Data analysis

## Dataset Description

The dataset that is being used has three major items: assigned insurance risk rating, normalized losses compared to other cars, and the various attributes for a car. There are a total of 26 attributes in the data set but we will be mainly considering the following four: symbolling, normalized losses, horsepower, and weight. This dataset contains a total of 205 models of cars.

## Symboling

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

Symboling refers to the assigned insurance risk rating. This is a normalized figure against all other vehicles.



Figure : Summary of Symboling

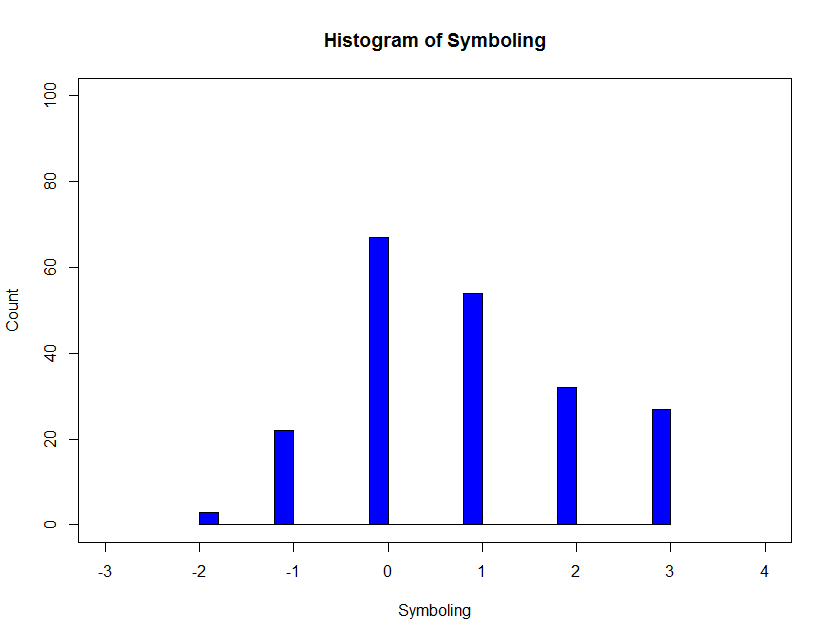


Figure : Histogram of Symboling

## Normalized Losses

Normalized losses is in comparison to other cars.



Figure : Normalized Losses

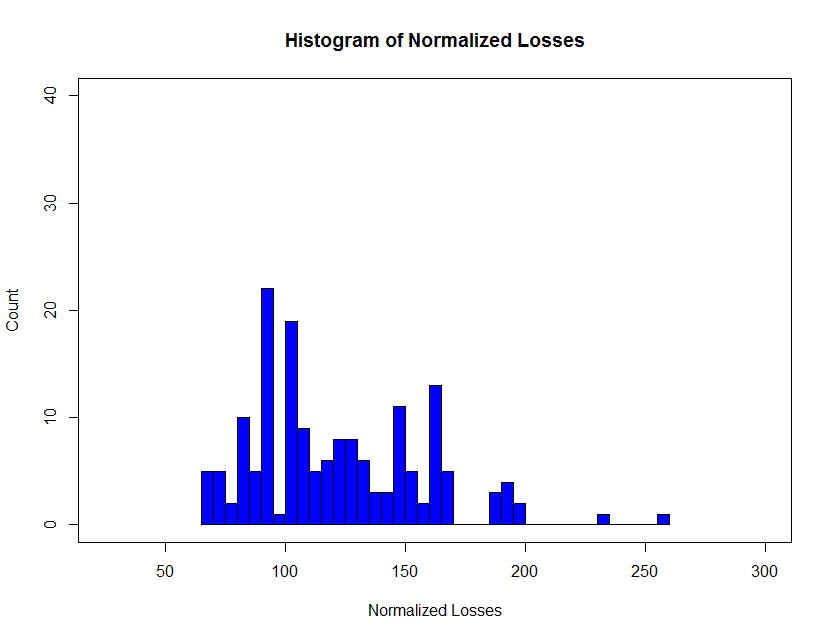


Figure : Histogram of Normalized Losses

## Horsepower

Horsepower refers to the rate at which work is done by the engine.



Figure : Horsepower

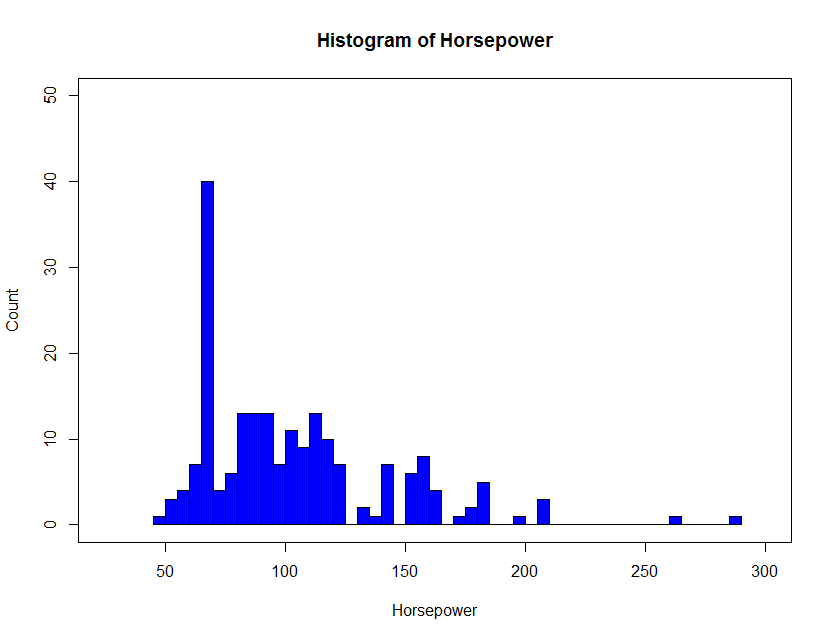


Figure : Histogram of Horsepower

## Weight

Weight refers to the vehicle curb weight in English pounds.



Figure : Weight

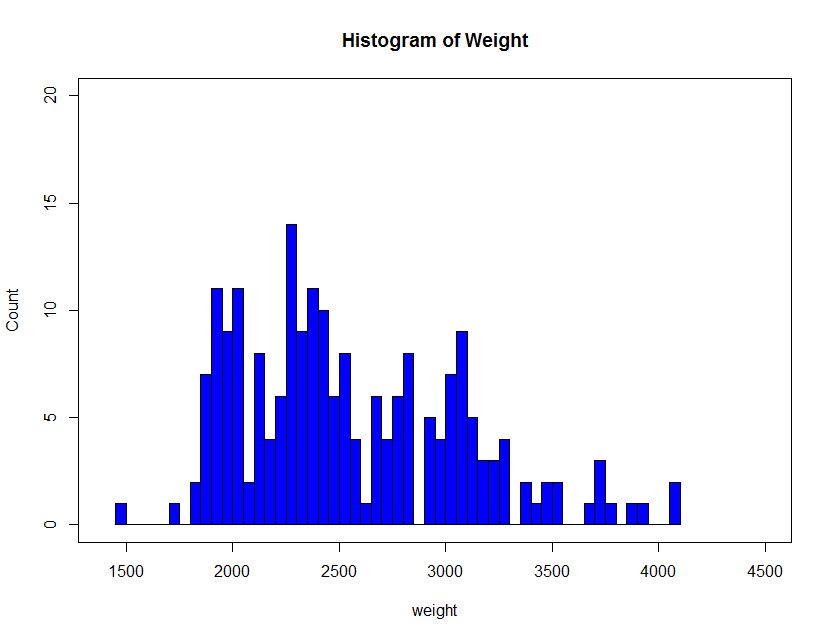


Figure : Histogram of Weight

## Weight vs. Horsepower

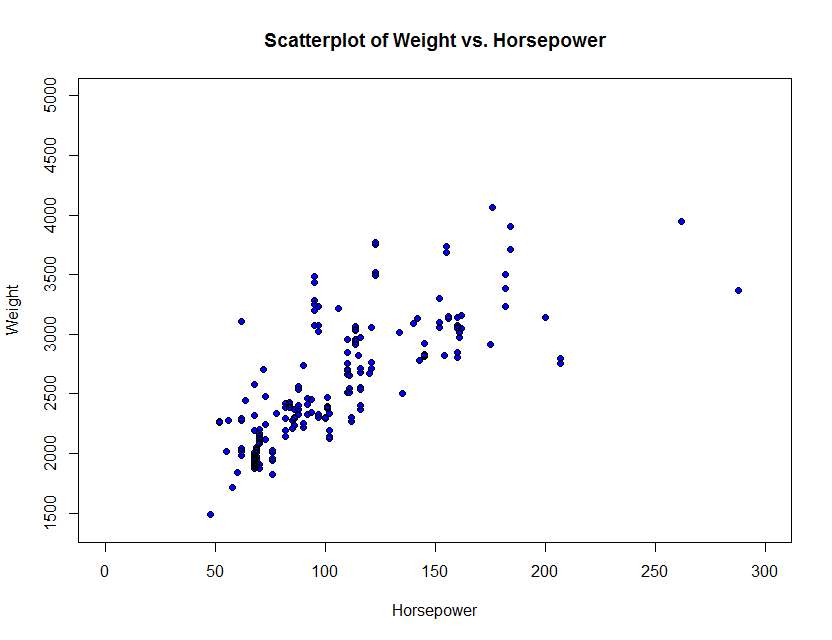


Figure : Weight vs. Horsepower

Weight and horsepower have a positive correlation.

## Symboling vs. Weight/Horsepower

Below features a scatterplot of symbolling (assigned insurance risk) and weight/horsepower.

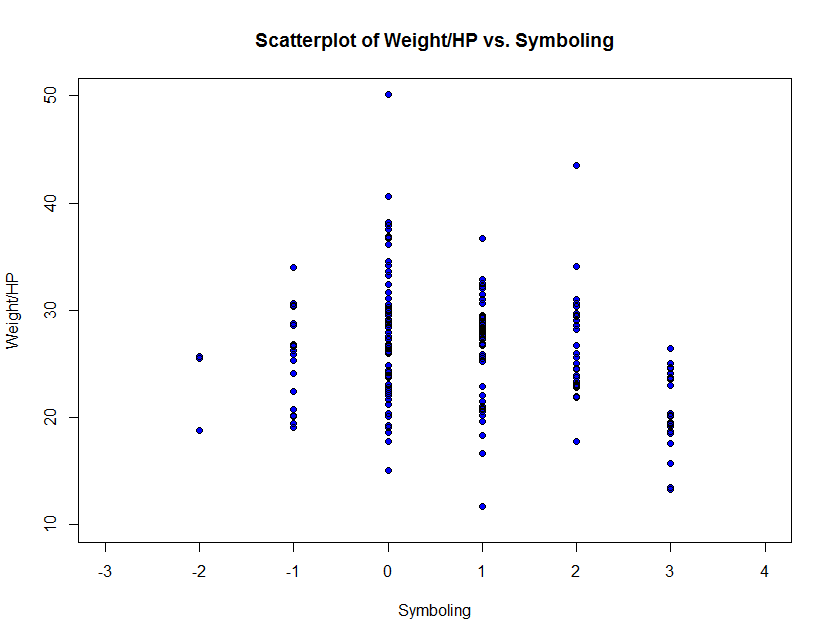


Figure : Scatterplot of Weight/HP vs. Symboling

## Normalized Losses vs. Weight/Horsepower

Below features a scatterplot of normalized losses and weight/horsepower.

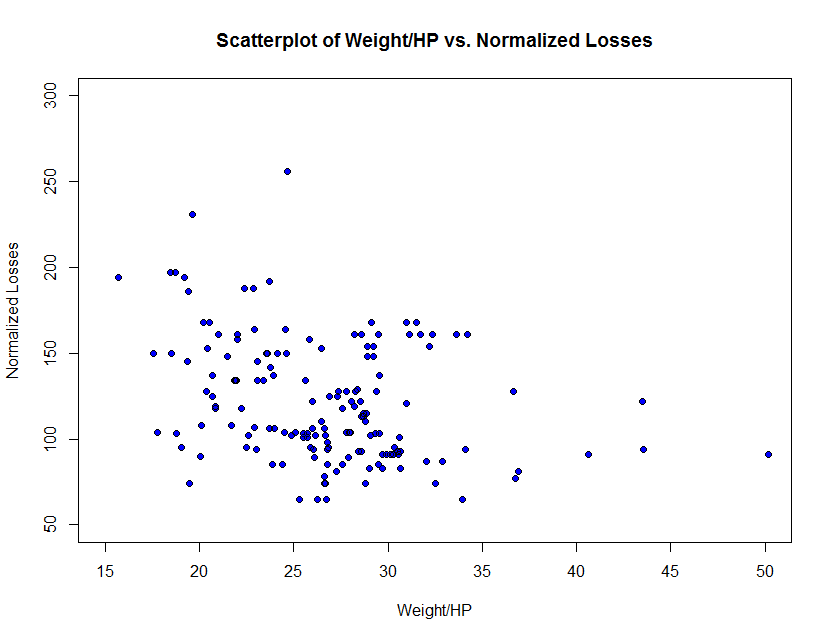


Figure : Scatterplot of Weight/HP vs. Normalized Losses

# Data preprocessing

## Normalizing “Horsepower” Data

### Min-Max Normalization

The horsepower data was normalized by first removing all the non-entries and then converting the entire column to a min-max normalized value. Below are the results with a min of 0 and max of 1.0.



Figure : Min-Max Normalization

### Z-Scores

The horsepower data was normalized by first removing all the non-entries and then converting the entire column to a Z-score value.



Figure : Z Score

### Decimal scaling

The horsepower data was normalized to values that go to the second decimal point.



Figure :Decimal scaling

## Binning

### Equal Frequency

Equal frequency binning divides the entire data set into three equal count bins.

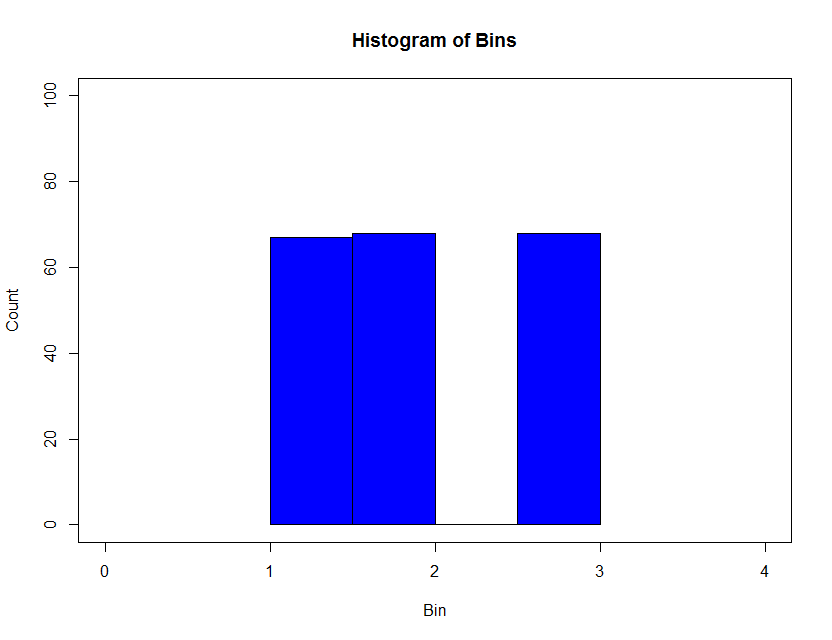
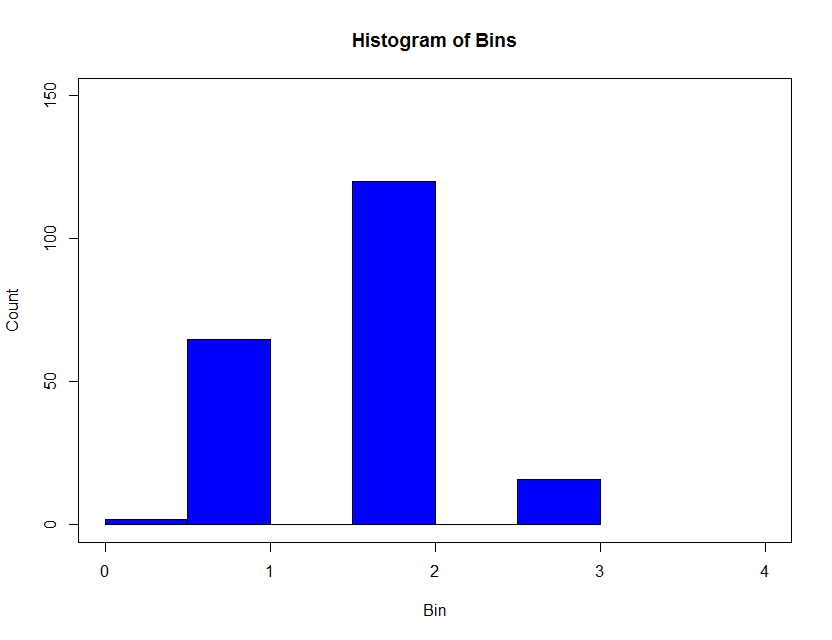


Figure : Histogram of Equal Frequency Binning

### Equal Width Binning

Equal width binning divides the range of the data set into n number of bins and results in a more normal distribution appearance in the histogram. This type of binning was more useful in separating the data than the previous



## Transformations

### Natural Log

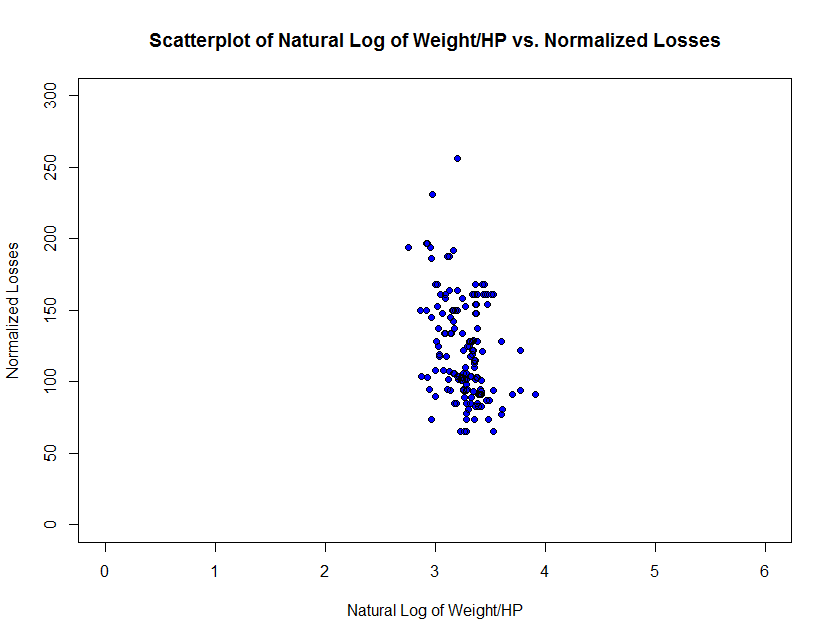


Figure : Scatterplot of Natural Log of Weight/HP vs. Normalized Losses

### Square Root

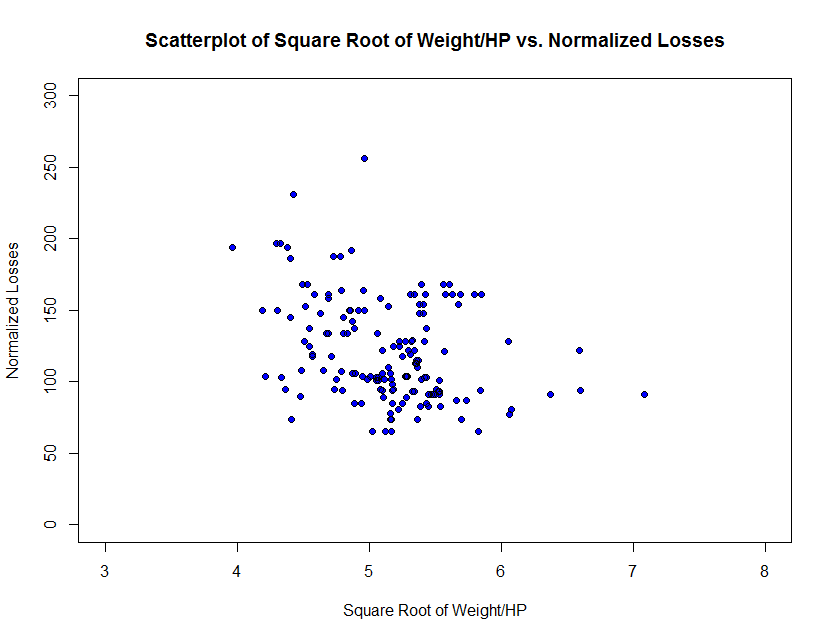


Figure : Scatterplot of Square Root of Weight/HP vs. Normalized Losses

### Inverse Square Root

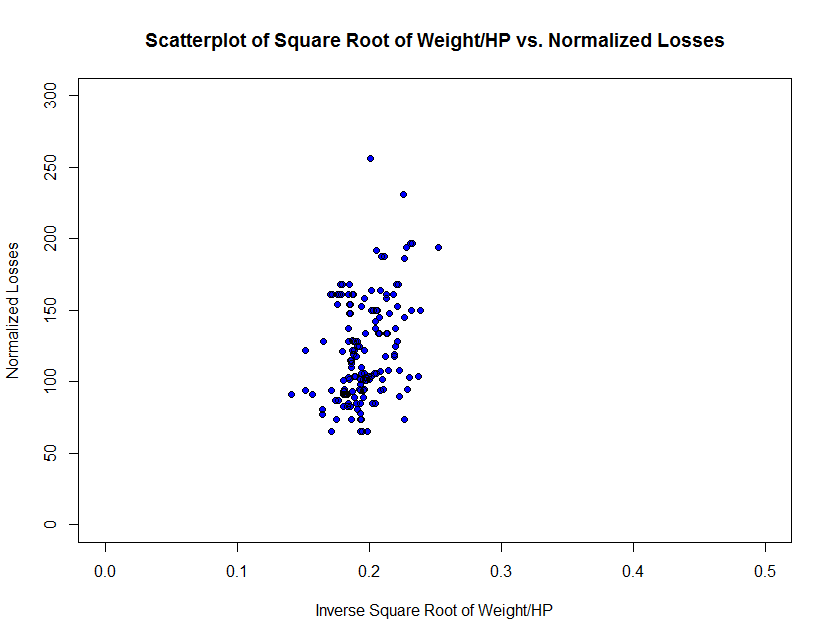


Figure : Scatterplot Inverse Square Root of Weight/HP vs. Normalized Losses

# regression analysis

## Prediction Question

The prediction question is does a low weight to horsepower (fast acceleration) indicate a car is more likely to wreck (normalized losses).

## Regression Model

### R Squared

The calculated fit model did have a low r-squared value which indicates that there are many other factors that are at play than simply weight to horsepower ratios when looking at normalized losses. This may indicate that normalized losses includes but is not limited to wrecks. It could involve losses from theft, natural disasters, etc.

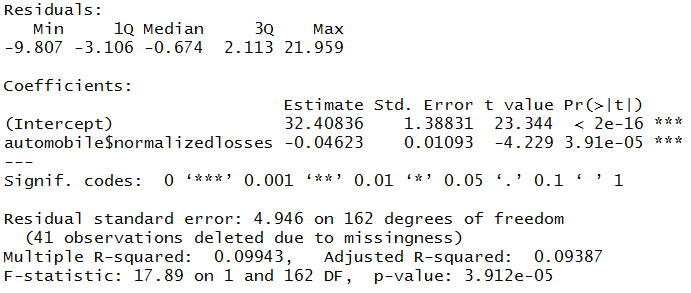


Figure : R2 Output

### P-value

We determined a very low p-value for the comparison between weight/horsepower and normalized losses. This indicates that the null hypothesis: sports cars aren’t more likely to be lost (or wrecked) is true.

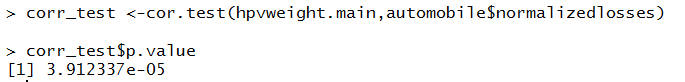


Figure : P-value