Car_Linear_Regression.R

lucasvalpreda

2025-08-06

```r
# Simple Linear Regression Project in R
# Dataset: mtcars (built-in)
# Purpose: Predict miles per gallon (mpg) from horsepower (hp)
# Author: Lucas Valpreda
```

```r
# loading the data set
data(mtcars)
# viewing the first rows of data
head(mtcars)
```

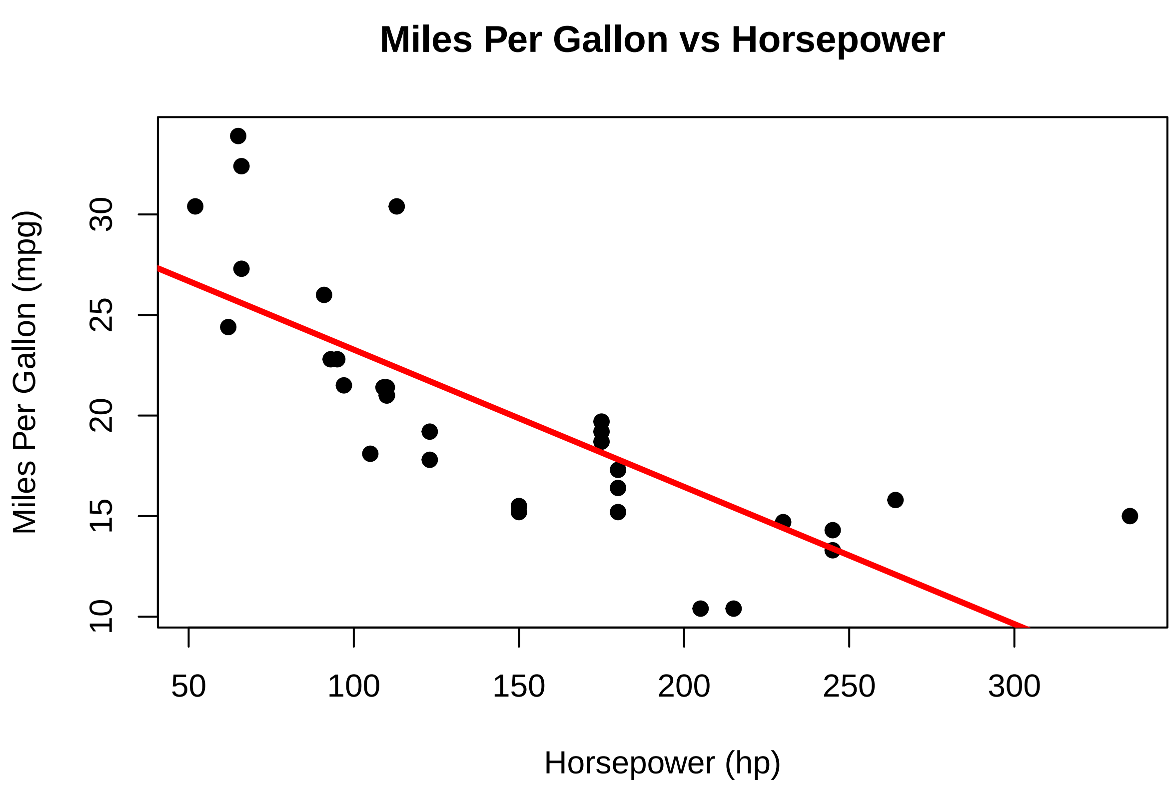
```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```r
# creating a scatter plot: mpg vs hp
plot(mtcars$hp, mtcars$mpg,
     main = "Miles Per Gallon vs Horsepower",
     xlab = "Horsepower (hp)",
     ylab = "Miles Per Gallon (mpg)",
     pch = 19,
     col= "Black"
     )
# now lets build a linear regression to determine if there is a relationship -
# between our dependent variable (mpg) vs our independent variable (hp)
model <- lm(mpg ~ hp, data = mtcars)
# lets view the model summary
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7121 -2.1122 -0.8854  1.5819  8.2360
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
## hp          -0.06823    0.01012  -6.742 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.863 on 30 degrees of freedom
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

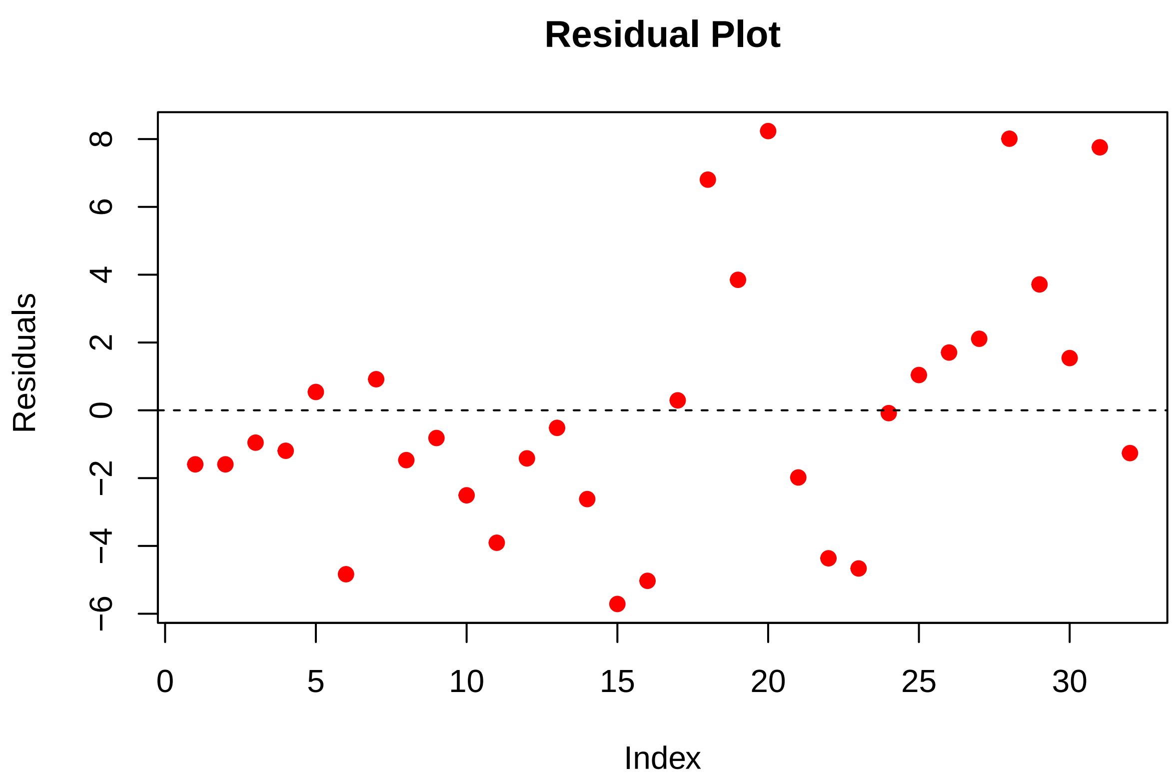
```r
# Plotting the data again
plot(mtcars$hp, mtcars$mpg,
     main = "Miles Per Gallon vs Horsepower",
     xlab = "Horsepower (hp)",
     ylab = "Miles Per Gallon (mpg)",
     pch = 19,
     col= "Black"
     )
# Adding the regression line
abline(model, col = "Red", lwd = 3)
```




```r
# lets predict the MPG for a new car with 150 hp
# creating a new data frame with hp = 150
new_data <- data.frame(hp = 150)
# now lets predict using our current model
predict(model, newdata = new_data)
```

```
##        1
## 19.86462
```

```r
# lets check residuals
plot(model$residuals,
     main = "Residual Plot",
     ylab = " Residuals",
     xlab = "Index",
     pch = 19,
     col = "red"
     )
abline(h = 0, lty =2)
```




```r
# Conclusion:
# This simple linear regression analysis used the built-in mtcars dataset
# to predict Miles per Gallon (mpg) based on Horsepower (hp).
#
# Key findings:
# - Regression equation: mpg = 30.10 - 0.068 * hp
# - R-squared: 0.60 → Horsepower explains 60% of the variance in mpg
# - P-value for hp: < 0.001 = Statistically significant
# - Residual plot suggests a decent linear fit with a few outliers
#
# Next steps (if extended):
# - Try multiple regression by adding more variables (e.g. weight)
# - Check for interaction effects or nonlinear models
```