

Linear Regression

Ravi Kumar Tiwari

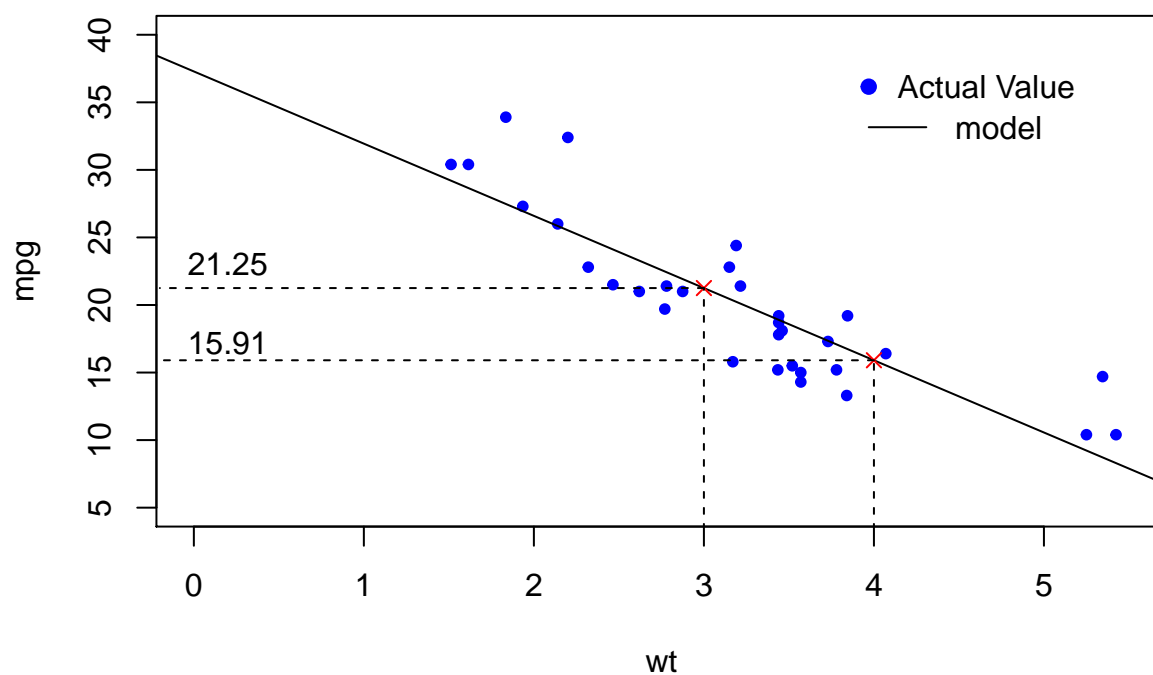
13 July 2016

Linear Regression

Linear Regression is used for modeling quantitative response. This method involves finding a straight line ($y = ax + b$) that best describes the relationship between the dependent and the independent variables. This best fit line is then used to predict the value of the dependent variable for any given values of the independent variables.

1: Example

Linear model depicting the relationship between wt (weight of the car) and mpg (miles per gallon) variable of mtcars datasets



2: Code for model building and prediction

Model Building

```
lmModel <- lm(mpg ~ wt, data = mtcars)
```

Model Prediction

```
predValue <- predict(lmModel, data.frame(wt = 3))  
predValue <- predict(lmModel, data.frame(wt = c(3,4)))  
predValue <- predict(lmModel, data.frame(wt = mtcars$wt))
```

3: Code to access the model parameters

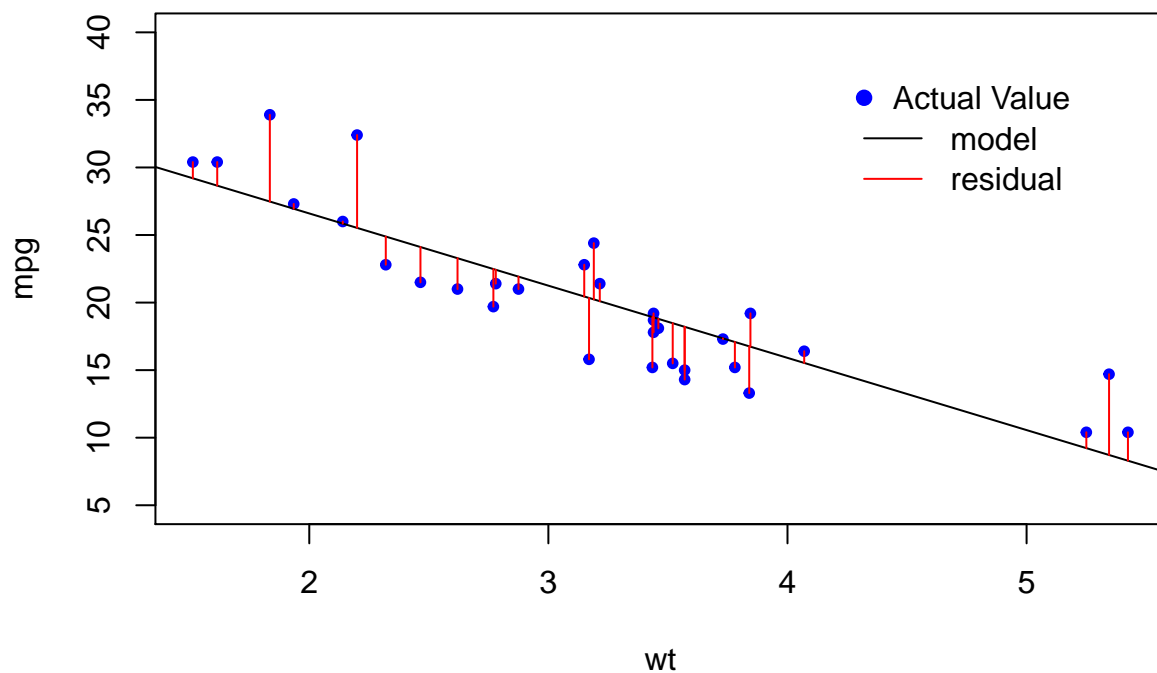
```
coef(lmModel)
```

```
## (Intercept)          wt  
##   37.285126    -5.344472
```

Interpretation: The intercept is the model prediction for the case when the independent variable is 0. The slope is the change in the dependent variable when the independent variable changes by 1 unit. Look at the plot, we created earlier and verify this is indeed true

4: Model performance assessment

I. Visual Inspection (Residual Plot)



II. R-squared value

R-squared is the measure of goodness of fit. Higher the R-squared value, better is the fit. Mathematically, $R\text{-squared} = \text{Explained Variation} / \text{Total Variation}$. In other words, R-squared is the percentage of the total variation in the y value that the model is able to explain

```
sumModel <- summary(lmModel)
sumModel$r.squared
```

```
## [1] 0.7528328
```

```
r <- cor(mtcars$mpg, mtcars$wt)
r^2
```

```
## [1] 0.7528328
```

r.squared is literally the r (correlation coefficient) squared

5: Code to build models with more than one predictors

```
lmModel2 <- lm(mpg ~ wt + hp + disp, data = mtcars) # wt, hp, and disp will be used as predictor
lmModel3 <- lm(mpg ~ ., data = mtcars) # All the variables will be used
```