1 Prediction Interval for Linear Regression

Assume that the error term ϵ in the simple linear regression model is independent of x, and is normally distributed, with zero mean and constant variance. For a given value of x, the interval estimate of the dependent variable y is called the prediction interval.

1.1 Problem

In the data set ***mtcars***, develop a 95% prediction interval of the mpg (miles per gallon) for the weight (wt) of 4.4, 4.8 and 4.8 tonnes.

Solution

We apply the `lm` function to a formula that describes the variable mpg by the variable wt, and save the linear regression model in a new variable `myModel`.

```
myModel <- lm(mpgs ~ wt, data = mtcars)
```

Then we create a new data frame called `myNewData`. This dataframe must have the same variable names of the explanatory variables used in the model fitting process.

```
myNewData <- data.frame(wt=c(4.4,4.6,4.8))
```

We now apply the `predict` function and set the predictor variable in the `newdata` argument. We also set the interval type as "`predict`", and use the default 0.95 confidence level.

```
> predict(myModel, newdata =myNewData,
    interval="predict")
  fit      lwr      upr
1 13.76945 7.309732 20.22917
2 12.70056 6.189263 19.21185
3 11.63166 5.061258 18.20207
>
```

Answer: The 95% prediction interval of the mpg (miles per gallon) for the wt time of 4.4 tonnes is between 7.31 and 20.22 tonnes.

Further detail of the `predict` function for linear regression model can be found in the R documentation (i.e. `help(predict.lm)`)