Week 2 Workshop

Dr Liwan Liyanage School of Computing Engineering and Mathematics Western Sydney University

28 July 2016

Data Upload

```r
library(MASS)
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 3.2.3
```

```r
attach(Boston)
```

Simple Linear Regression

- ▶ Data Understanding

```
names(Boston)
```

```
## [1] "crim"    "zn"      "indus"   "chas"    "nox"
## [8] "dis"     "rad"     "tax"     "ptratio" "black"
```

```
head(Boston)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222
##   lstat medv
## 1  4.98 24.0
```
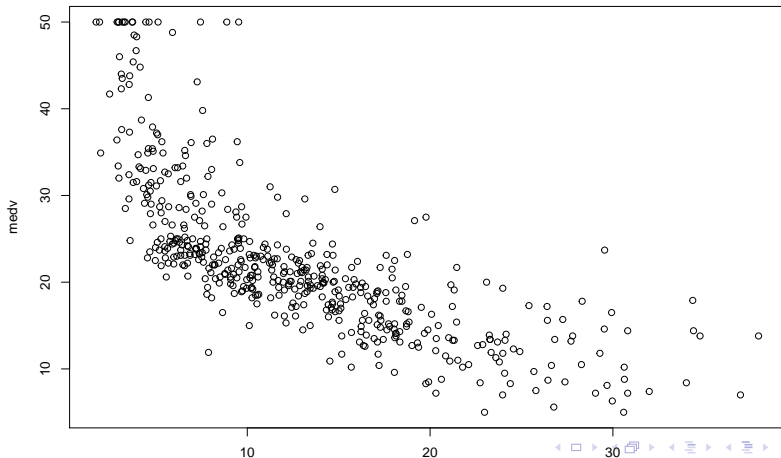
Explore details of the Boston data set

```
?Boston
```

```
## starting httpd help server ... done
```

What is the Target Variable?

  ▶ scatter plot

```
plot(medv~lstat,Boston)
```

Simple linear regression model

```
model1 = lm(medv~lstat)
model1
```

```
##
## Call:
## lm(formula = medv ~ lstat)
##
## Coefficients:
## (Intercept)        lstat
##       34.55        -0.95
```

```
summary(model1)
```

```
##
## Call:
## lm(formula = medv ~ lstat)
##
## Residuals:
##     Min     1Q  Median     3Q    Max
## -15.168 -3.990 -1.318  2.034 24.500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.55384    0.56263   61.41   <2e-16 ***
## lstat       -0.95005    0.03873  -24.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 6.216 on 504 degrees of freedom
## Multiple R-squared:  0.5441, Adjusted R-squared:  0.5432
```
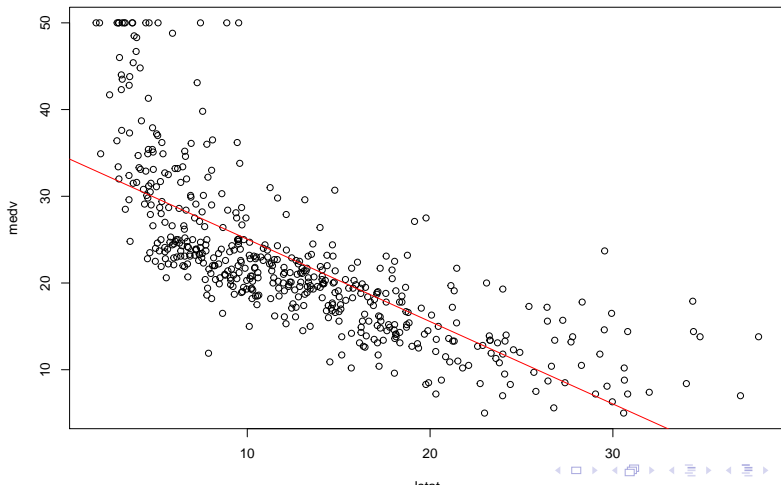
Plot the linear model within the scatterplot

```
plot(medv~lstat,Boston)
abline(model1,col="red")
```

```r
names(model1)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "ra
##  [5] "fitted.values" "assign"        "qr"            "df
##  [9] "xlevels"       "call"          "terms"         "mo
```

Coefficients and Predicted values

```
confint(model1)
```

```
##                   2.5 %      97.5 %
## (Intercept) 33.448457 35.6592247
## lstat       -1.026148 -0.8739505
```
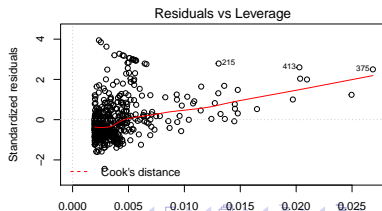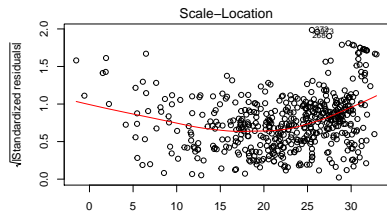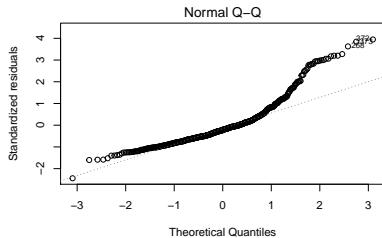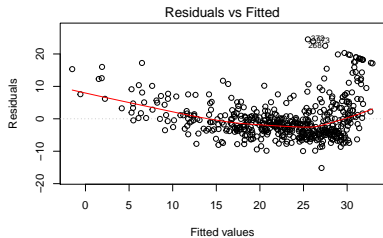
```
predict(model1,data.frame(lstat= c(5,10,15,20)),interval="
```

```
##        fit      lwr      upr
## 1 29.80359 29.00741 30.59978
## 2 25.05335 24.47413 25.63256
## 3 20.30310 19.73159 20.87461
## 4 15.55285 14.77355 16.33216
```

NOTE: We could also use list rather than data.frame such as -
predict(model1,list(lstat= c(5,10,15,20)),interval="confidence")

Model Checking

```
par(mfrow=c(2,2))
plot(model1)
```

Exercises

Please upload the following via Turnitin link given in vUWS site.

Construct the linear Regression models for medv against all other numerical predictor variables.

- ▶ 1 Select all predictor variables having a significant linear relationship with the target variable medv.
- ▶ 2 For each linear model give the regression equation and its R squared value and discuss its strength.
- ▶ 3 plot the scater plot with the abline for all significant linear models