Experiment 3

Prof Meenakshi Garg

7/31/2020

Regression Analysis

Implementation and analysis of Linear regression through graphical methods including Plots

setwd("E:/R Orientation")

```
library(ggplot2)
my_data <-mtcars
names(my_data)
```

[1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear" ## [11] "carb"

```
dim(my_data)
```

[1] 32 11

```
#randomize
my_data <- my_data[sample(nrow(my_data), ), ]
head(my_data)
```

mpg cyl disp hp drat wt qsec vs am gear carb ## Mazda RX4 Wag 21.0 6 160.0 110 3.90 2.875 17.02 0 1 4 4 ## Duster 360 14.3 8 360.0 245 3.21 3.570 15.84 0 0 3 4 ## Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.50 0 1 5 4 ## Mazda RX4 21.0 6 160.0 110 3.90 2.620 16.46 0 1 4 4 ## Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.70 0 1 5 2 ## Cadillac Fleetwood 10.4 8 472.0 205 2.93 5.250 17.98 0 0 3 4

```
TrainData <- my_data[1:20,]
TestData <- my_data[21:32,]
## Linear Model
fit = lm(mpg ~ hp, data=mtcars)
summary(fit)
```

```
##
## 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
```

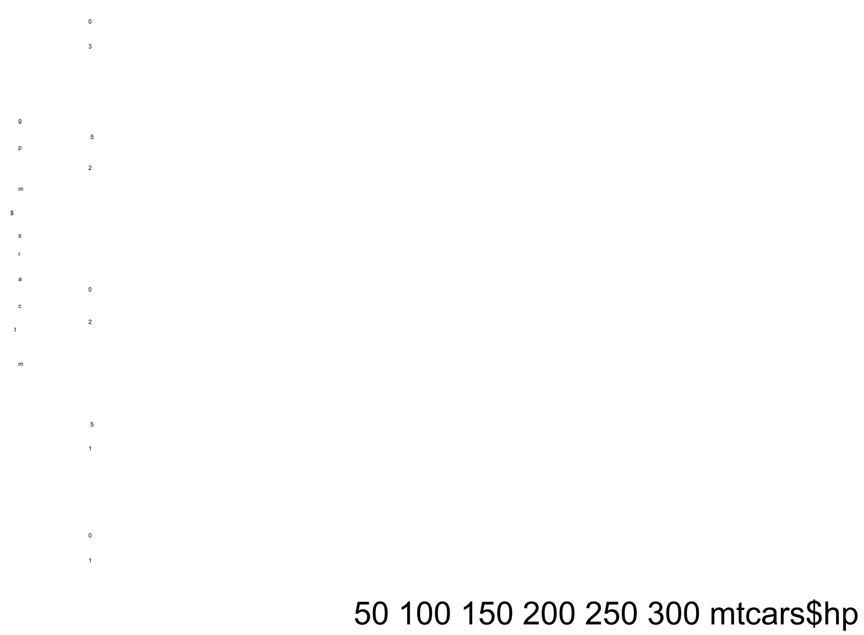
```
preds <- predict(fit, newdata = TestData)
df1 <- data.frame(preds,TestData$mpg)
head(df1)
```

```
## preds TestData.mpg
## Hornet 4 Drive 22.59375 21.4
## Chrysler Imperial 14.40636 14.7
## Merc 230 23.61717 22.8
## Volvo 142E 22.66198 21.4
## Datsun 710 23.75363 22.8
## Camaro Z28 13.38293 13.3
```

```
#correlation
cor(preds,TestData$mpg)
```

```
## [1] 0.7858354
```

```
plot(mtcars$hp, mtcars$mpg)
```




```
ggplot(fit, aes(hp, mpg)) +
geom_point() +
stat_smooth(method = lm, se = FALSE) +
  geom_segment(aes(xend = hp, yend = .fitted), color = "red", size = 0.3)
```

20

15

10

35

100 200 300 hp

30

25

```
# BetterModel ?
lmmodel1 <- lm(mpg ~ hp+cyl+gear+wt, data = TrainData)
summary(lmmodel1)
```

```
##
## Call:
## lm(formula = mpg ~ hp + cyl + gear + wt, data = TrainData) ##
## Residuals:
## Min 1Q Median 3Q Max
## -3.9265 -1.5118 -0.6719 1.2700 5.3709
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|) ## (Intercept) 39.28913
7.48216 5.251 9.77e-05 *** ## hp -0.02878 0.02567 -1.122
0.2797 ## cyl -0.70088 1.05380 -0.665 0.5161 ## gear 0.03884
1.22367 0.032 0.9751 ## wt -3.40102 1.26458 -2.689 0.0168 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ##
## Residual standard error: 2.784 on 15 degrees of freedom ##
Multiple R-squared: 0.8446, Adjusted R-squared: 0.8031 ##
F-statistic: 20.38 on 4 and 15 DF, p-value: 6.332e-06
```



```
preds_new <- predict(lmmodel1, newdata = TestData)
df2 <- data.frame(preds_new,TestData$mpg)
head(df2)
```

```
## preds_new TestData.mpg
## Hornet 4 Drive 21.099790 21.4
```

```
## Chrysler Imperial 8.999695 14.7
## Merc 230 23.193219 22.8
## Volvo 142E 24.048611 21.4
## Datsun 710 26.073636 22.8
## Camaro Z28 13.686462 13.3
```

```r
cor(preds_new,TestData$mpg)
```

```
## [1] 0.9266731
```

```r
plot(mtcars$hp+mtcars$cyl+mtcars$gear+mtcars$wt, mtcars$mpg)
```


```r
ggplot(fit, aes(mtcars$hp+mtcars$cyl+mtcars$gear+mtcars$wt, mpg)) +
geom_point() +
stat_smooth(method = lm, se = FALSE) +
  geom_segment(aes(xend = hp, yend = .fitted), color = "red", size = 0.3)
```

mp

m
35

30

25

20

15

10

100 200 300 mtcars$hp + mtcars$cyl + mtcars$gear + mtcars$wt

Linear Regression