RegressionModels.R

dieudonneouedraogo

Thu Mar 2 13:35:06 2017

```
##############################################################################################

########THIS IS A R SCRIPT ,USE THE PROVIDED DATASETS TO RUN REGRESSION MODELS############

#################SSIE500/PROFESSOR HIROKI SAYAMA & DIEUDONNE OUEDRAOGO####################

##############################################################################################

#DATASET :data0
require(readr)|| install.packages("readr")
```

```
## Loading required package: readr
```

```
## [1] TRUE
```

```
require(knitr)|| install.packages("knitr")
```

```
## Loading required package: knitr
```

```
## [1] TRUE
```

```
data0 <- read_delim("~/data0.txt", " ", escape_double = FALSE)
kable(data0)
```

Z1	Z2	Z3	Z4	Z5	Y
0.375	3.13	60.0	40	2.00	101
1.000	3.13	76.8	30	1.99	141
1.000	3.13	60.0	20	2.00	96
1.000	3.13	60.0	20	1.98	125
1.625	3.13	43.2	10	2.01	43
1.625	3.13	60.0	20	2.00	16
1.625	3.13	60.0	20	2.02	188
0.375	5.00	76.8	10	2.01	10
1.000	5.00	43.2	10	1.99	3
1.000	5.00	43.2	30	2.01	386
1.000	5.00	100.0	20	2.00	45
1.625	5.00	76.8	10	1.99	2
0.375	1.25	76.8	10	2.01	76
1.000	1.25	43.2	10	1.99	78
1.000	1.25	76.8	30	2.00	160
1.000	1.25	60.0	0	2.00	3
1.625	1.25	43.2	30	1.99	216
1.625	1.25	60.0	20	2.00	73
0.375	3.13	76.8	30	1.99	314
0.375	3.13	60.0	20	2.00	170

```
fit0=lm(log(Y)~Z1+Z2+Z3+Z4+Z5,data=data0)#Linear model log(y) versus other components
# Stepwise Regression
```

```r
require(MASS)|| install.packages(MASS)
```

```
## Loading required package: MASS
```

```
## [1] TRUE
```

```r
step <- stepAIC(fit0, direction="both")#Stepwise Selection
```

```
## Start:  AIC=7.58
## log(Y) ~ Z1 + Z2 + Z3 + Z4 + Z5
##
##        Df Sum of Sq    RSS     AIC
## - Z3    1    0.4917 16.523  6.1810
## - Z1    1    0.8006 16.832  6.5514
## <none>              16.032  7.5768
## - Z5    1    1.9995 18.031  7.9275
## - Z2    1    3.9387 19.971  9.9705
## - Z4    1   24.5815 40.613 24.1673
##
## Step:  AIC=6.18
## log(Y) ~ Z1 + Z2 + Z4 + Z5
##
##        Df Sum of Sq    RSS     AIC
## - Z1    1    0.5160 17.039  4.7960
## <none>              16.523  6.1810
## - Z5    1    1.9690 18.492  6.4327
## + Z3    1    0.4917 16.032  7.5768
## - Z2    1    4.5731 21.097  9.0675
## - Z4    1   24.3922 40.916 22.3156
##
## Step:  AIC=4.8
## log(Y) ~ Z2 + Z4 + Z5
##
##        Df Sum of Sq    RSS     AIC
## <none>              17.039  4.7960
## - Z5    1    1.9747 19.014  4.9890
## + Z1    1    0.5160 16.523  6.1810
## + Z3    1    0.2071 16.832  6.5514
## - Z2    1    4.3410 21.380  7.3349
## - Z4    1   25.8384 42.878 21.2524
```

```r
step$anova # display results
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## log(Y) ~ Z1 + Z2 + Z3 + Z4 + Z5
##
## Final Model:
## log(Y) ~ Z2 + Z4 + Z5
##
##
##   Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                         14   16.03180 7.576834
## 2 - Z3  1 0.4916555        15   16.52345 6.180967
```

```
## 3 - Z1  1 0.5159839        16   17.03943 4.795961
```

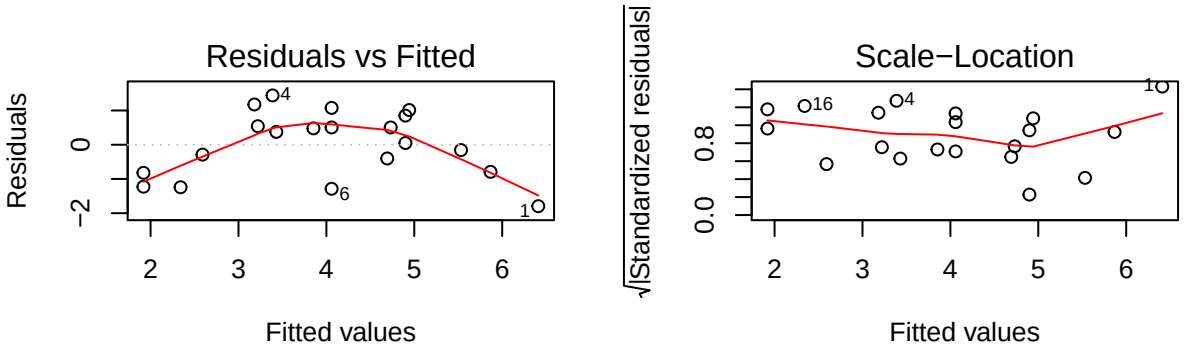
```
#plot(step$residuals)
coefficients(step) # model coefficients
```

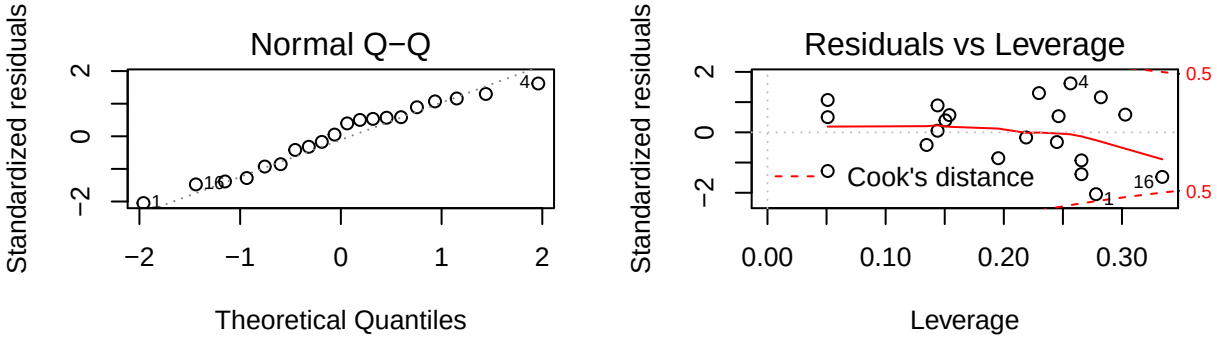
```
## (Intercept)          Z2          Z4          Z5
## -64.4321536  -0.3364702   0.1175412  33.5970841
```

```
anova(step) # anova table
```

```
## Analysis of Variance Table
##
## Response: log(Y)
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Z2         1  4.0442  4.0442  3.7975 0.0690868 .
## Z4         1 24.5512 24.5512 23.0536 0.0001958 ***
## Z5         1  1.9747  1.9747  1.8542 0.1921612
## Residuals 16 17.0394  1.0650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(step)
```





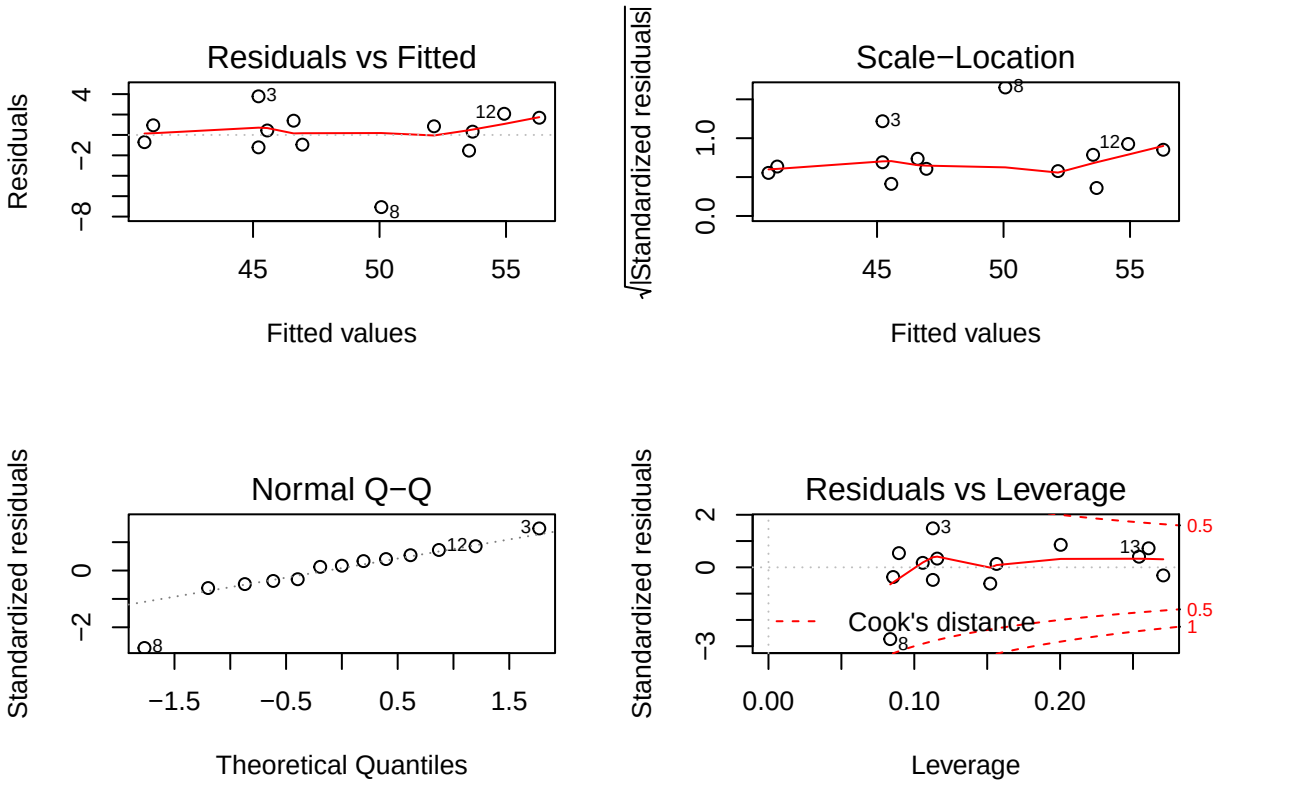

```
##############################################################################################

#DATASET :data2

require(readr)
data2 <- read_csv("~/Downloads/data2.csv")
kable(data2)
```

Y	X
40	825
42	830
49	890
46	895
44	890
48	910
46	915
43	960
53	990
52	1010
54	1012
57	1030
58	1050

```r
fit2=lm(Y~X,data=data2)#Linear Model Y versus X
summary(fit2)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0718 -0.9509  0.4363  1.3959  3.7830
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.50931    9.84346  -1.677    0.122
## X             0.06936    0.01045   6.635 3.68e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.706 on 11 degrees of freedom
## Multiple R-squared:  0.8001, Adjusted R-squared:  0.7819
## F-statistic: 44.03 on 1 and 11 DF,  p-value: 3.683e-05
```

```r
layout(matrix(c(1,2,3,4),2,2)) # optional 4 graphs/page
plot(fit2)
```


Residuals vs Fitted
Residuals
Fitted values
Scale−Location
√|Standardized residuals|
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

##