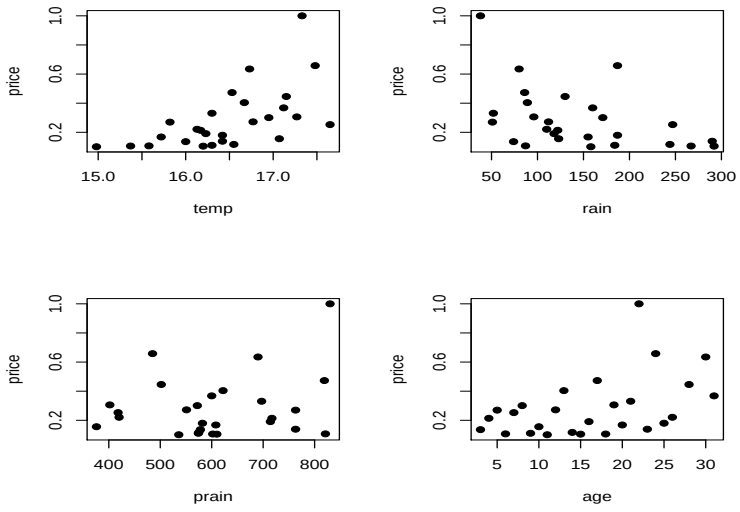
## Chapter 9: Model checking

```
> wine=read.table("wine.txt", header=T)[,-1]
> #attach(wine)
> cor(wine) #correlation coefficients
```

```
        price     temp     rain    prain      age
price  1.0000  0.58742 -0.44923  0.23401  0.45236
temp   0.5874  1.00000 -0.02615 -0.32193  0.29437
rain  -0.4492 -0.02615  1.00000 -0.26799  0.05885
prain  0.2340 -0.32193 -0.26799  1.00000 -0.05118
age    0.4524  0.29437  0.05885 -0.05118  1.00000
```

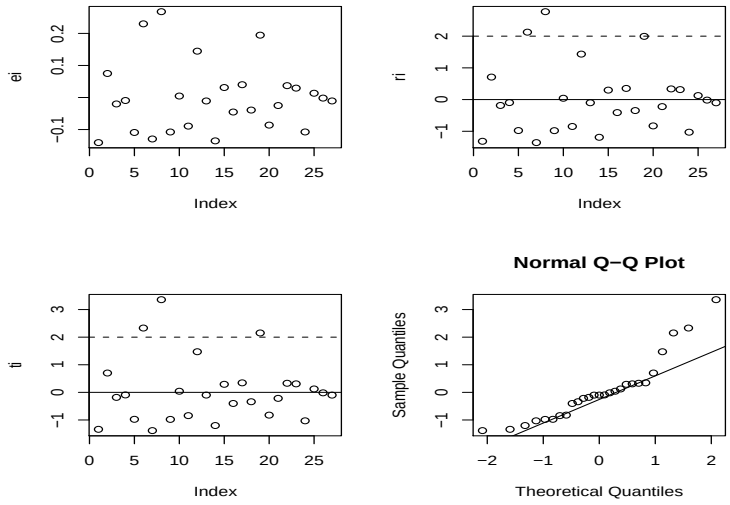
```
> #see data
> par(mfrow=c(2,2))
> plot(temp,price,pch=16)
> plot(rain,price,pch=16)
> plot(prain,price,pch=16)
> plot(age,price,pch=16)
```



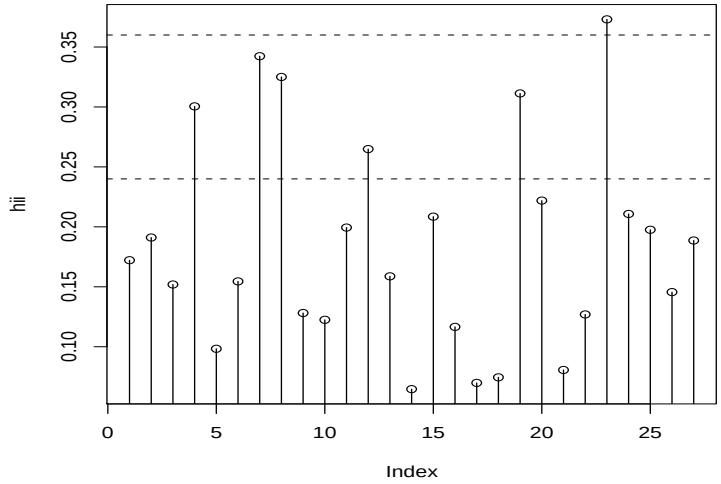

```
> f=lm(price~temp+rain+prain+age, data=wine)
> anova(f)
```

```
Analysis of Variance Table

Response: price
          Df Sum Sq Mean Sq F value  Pr(>F)
temp       1  0.397   0.397   28.71 2.2e-05
rain       1  0.217   0.217   15.67 0.00067
prain      1  0.129   0.129    9.32 0.00583
age        1  0.104   0.104    7.50 0.01201
Residuals 22  0.304   0.014
```

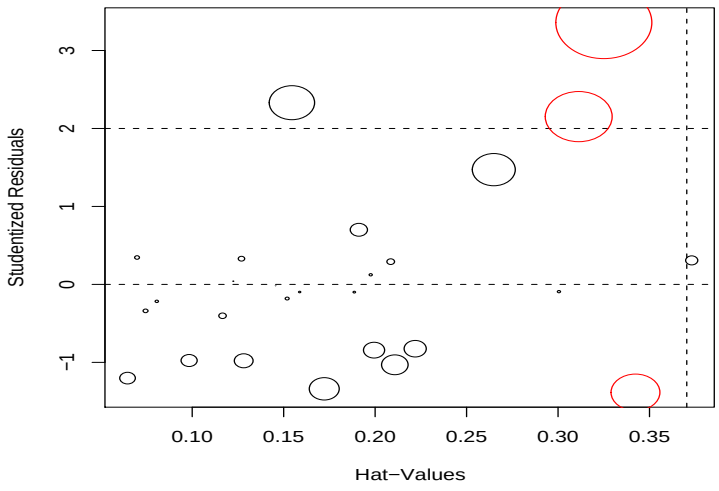
```
> summary(f)

Call:
lm(formula = price ~ temp + rain + prain + age, data = wine)

Residuals:
    Min      1Q  Median      3Q     Max
-0.1407 -0.0877 -0.0107  0.0341  0.2678

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.171629   0.692890   -4.58  0.00015
temp         0.190310   0.039061    4.87  7.2e-05
rain        -0.001035   0.000331   -3.12  0.00495
prain        0.000564   0.000198    2.85  0.00934
age          0.008052   0.002941    2.74  0.01201

Residual standard error: 0.118 on 22 degrees of freedom
Multiple R-squared: 0.736,        Adjusted R-squared: 0.687
F-statistic: 15.3 on 4 and 22 DF,  p-value: 4.02e-06

> y.hat=fitted(f)
> ei=residuals(f)  #residual: y-y.hat
> hii=hatvalues(f)
> deleted.ei=ei/(1-hii) #deleted residual (press residual): y-y.hat(i)
> ri=rstandard(f)  #ei/(S*sqrt(1-hii)) studentized residual, standardized press
> ti=rstudent(f)   #delete.ei/(S(i)*sqrt(1-hii))
> (press=sum(deleted.ei^2))

[1] 0.5341

> par(mfrow=c(2,2))
> plot(ei)
> plot(ri);abline(h=c(0,-2,2), lty=c(1,2,2))
> plot(ti);abline(h=c(0,-2,2), lty=c(1,2,2))
> qqnorm(ti); qqline(ti) #qq.plot(f, simulate=T)  #identify few possible outliers
```



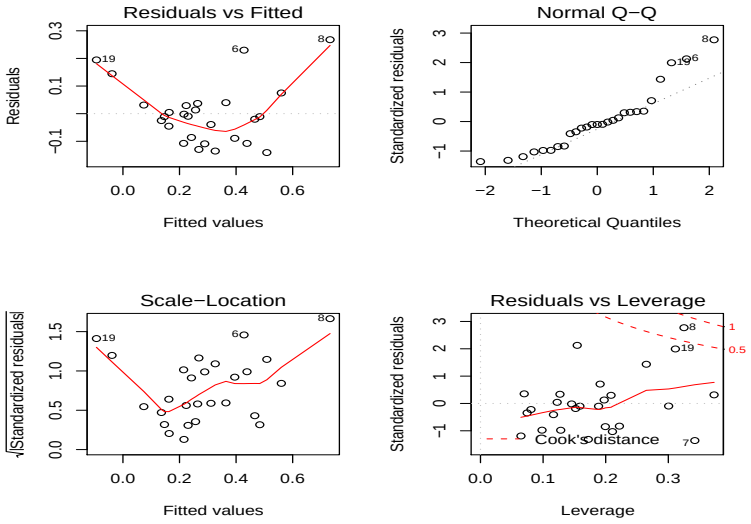

```
> par(mfrow=c(1,1))
> plot(hii, type="h")  #type for vertical line
> points(hii)
> abline(h=c(2, 3)*3/25, lty=2)
> n=nrow(wine) #number of observations
> identify(1:n, hii) #identify

integer(0)
```



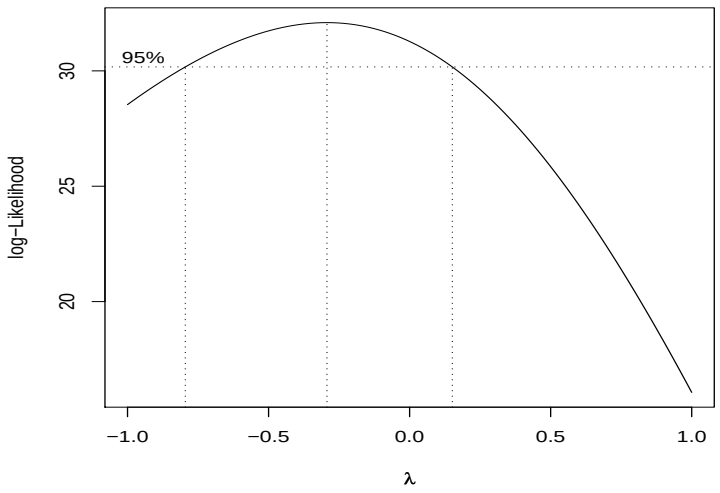

```
> library(car)
> influencePlot(f)

integer(0)
```



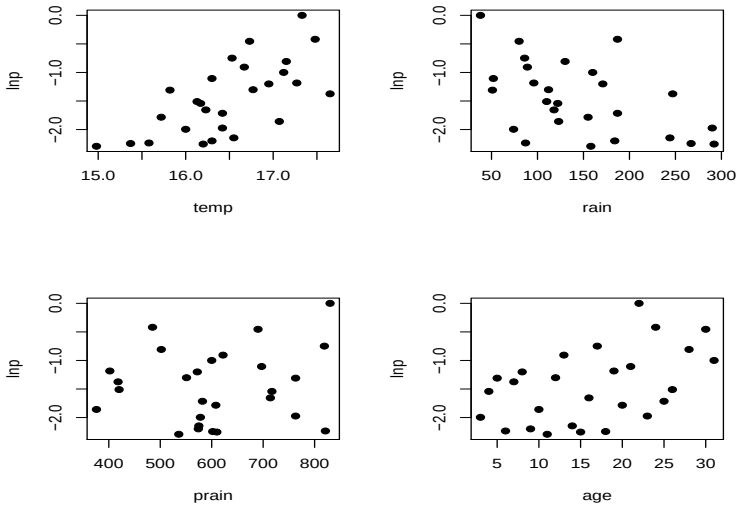

```
> par(mfrow=c(2,2))
> plot(f, ask=F)
```




```
> library(MASS)
> par(mfrow=c(1,1))
> boxcox(f, plotit=T, lambda=seq(-1,1,by=0.1))
```




```
> lnp=log(price)
> par(mfrow=c(2,2))
> plot(temp,lnp,pch=16)
> plot(rain,lnp,pch=16)
> plot(prain,lnp,pch=16)
> plot(age,lnp,pch=16)
```




```
> f1=lm(lnp~temp+rain+prain+age, data=wine)
> summary(f1)

Call:
lm(formula = lnp ~ temp + rain + prain + age, data = wine)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-0.4575 -0.2390  0.0107  0.1853  0.5364

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.22e+01   1.69e+00   -7.21  3.1e-07
temp         6.17e-01   9.50e-02    6.49  1.6e-06
rain        -3.87e-03   8.06e-04   -4.80  8.7e-05
prain        1.17e-03   4.81e-04    2.43   0.0236
age          2.39e-02   7.15e-03    3.34   0.0030

Residual standard error: 0.286 on 22 degrees of freedom
Multiple R-squared: 0.828,        Adjusted R-squared: 0.797
F-statistic: 26.5 on 4 and 22 DF,  p-value: 3.89e-08
```

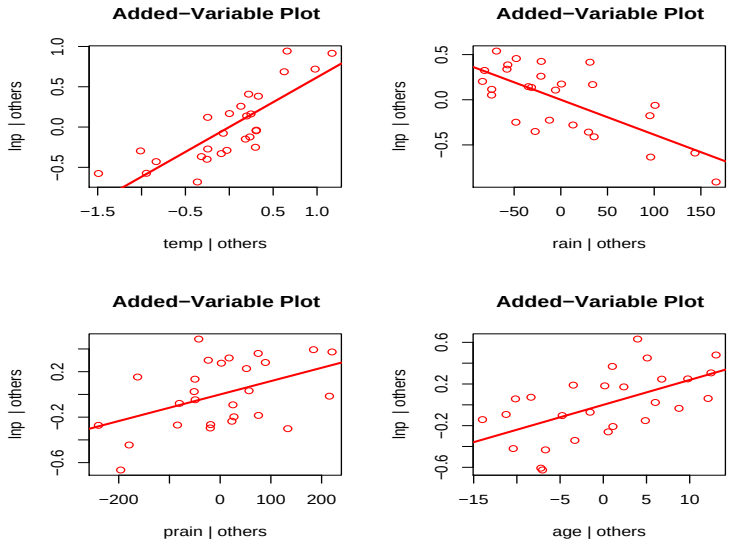
```
> anova(f1)

Analysis of Variance Table

Response: lnp
          Df Sum Sq Mean Sq F value  Pr(>F)
temp       1   4.66    4.66   56.99 1.5e-07
rain       1   2.52    2.52   30.73 1.4e-05
prain      1   0.59    0.59    7.18  0.0137
age        1   0.91    0.91   11.16  0.0030
Residuals 22   1.80    0.08
```

```
> confint(f1)

                 2.5 %    97.5 %
(Intercept) -1.565e+01 -8.662831
temp         4.199e-01  0.814052
rain        -5.538e-03 -0.002194
prain        1.726e-04  0.002169
age          9.064e-03  0.038739
```

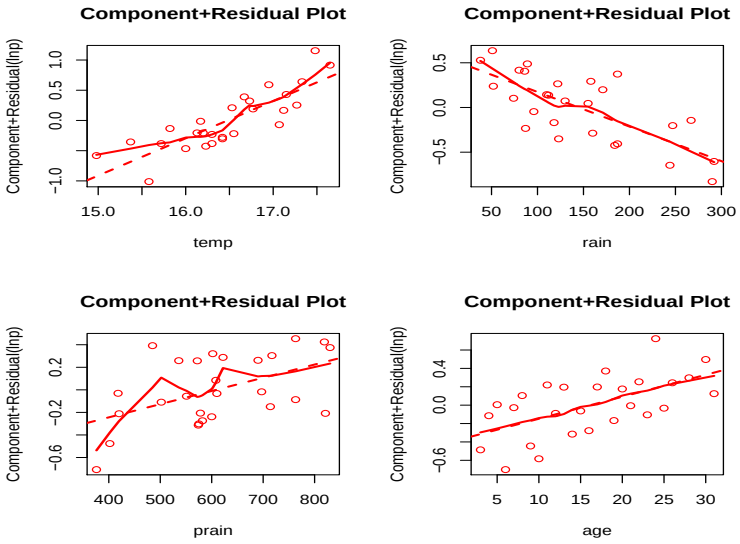
```
> a=predict(f1, interval="predict")
> cbind(price, exp(a))

   price     fit     lwr    upr
1  0.368 0.46267 0.24340 0.8795
2  0.635 0.53762 0.28137 1.0272
3  0.446 0.43924 0.23236 0.8303
4  0.221 0.21903 0.11135 0.4309
5  0.180 0.22958 0.12329 0.4275
6  0.658 0.38482 0.20343 0.7280
7  0.139 0.18167 0.09136 0.3613
8  1.000 0.89108 0.45012 1.7640
9  0.331 0.37359 0.19894 0.7016
10 0.168 0.15432 0.08231 0.2893
11 0.306 0.38696 0.20206 0.7410
```

```
12 0.106 0.07635 0.03917 0.1488
13 0.473 0.39579 0.20898 0.7496
14 0.191 0.25094 0.13606 0.4628
15 0.105 0.10868 0.05661 0.2086
16 0.117 0.15217 0.08130 0.2848
17 0.404 0.30778 0.16662 0.5685
18 0.272 0.26911 0.14550 0.4978
19 0.101 0.07162 0.03631 0.1413
20 0.156 0.24104 0.12511 0.4644
21 0.111 0.14577 0.07867 0.2701
22 0.301 0.22298 0.11878 0.4186
23 0.253 0.20871 0.10414 0.4183
24 0.107 0.16907 0.08801 0.3248
25 0.270 0.20556 0.10739 0.3935
26 0.214 0.17937 0.09505 0.3385
27 0.136 0.16133 0.08449 0.3081
```

```
> par(mfrow=c(2,2))
> av.plot(f1, temp, identify=F)
> av.plot(f1, rain, identify=F)
> av.plot(f1, prain, identify=F)
> av.plot(f1, age, identify=F)
```




```
> par(mfrow=c(2,2))
> cr.plot(f1, temp)
> cr.plot(f1, rain)
> cr.plot(f1, prain)
> cr.plot(f1, age)
```




```
> #par(mfrow=c(1,1))
> influence.measures(f1)

Influence measures of
         lm(formula = lnp ~ temp + rain + prain + age, data = wine) :

     dfb.1_  dfb.temp dfb.rain dfb.pran  dfb.age   dffit cov.r   cook.d
1   0.11558 -0.098487 -0.03532 -0.04403 -0.29166 -0.3990 1.275 3.22e-02
2  -0.01471  0.000391 -0.11371  0.06197  0.23161  0.3102 1.417 1.98e-02
3  -0.00258  0.004048 -0.00546 -0.00833  0.01504  0.0240 1.487 1.20e-04
4   0.01584 -0.013712 -0.01006 -0.01765  0.01363  0.0240 1.803 1.20e-04
5  -0.06502  0.079137 -0.07850  0.02977 -0.20566 -0.2944 1.162 1.75e-02
6  -0.47526  0.526482  0.21163 -0.17127  0.23901  0.9458 0.524 1.52e-01
7   0.24385 -0.111902 -0.67217 -0.52004 -0.17463 -0.8393 1.404 1.39e-01
8  -0.20708  0.190675 -0.09541  0.21783  0.03491  0.3345 1.769 2.32e-02
9  -0.03842  0.038816  0.11264 -0.02020 -0.07171 -0.1705 1.381 6.04e-03
10  0.08877 -0.091719 -0.00530 -0.02859  0.05603  0.1160 1.405 2.80e-03
11  0.02359 -0.103803  0.21337  0.30843 -0.02847 -0.4560 1.297 4.19e-02
12  0.47235 -0.534676  0.45231 -0.06278  0.20244  0.8187 1.123 1.29e-01
13 -0.10668  0.080460 -0.04387  0.22648  0.00311  0.2913 1.349 1.74e-02
14 -0.01404  0.030904  0.03498 -0.12954 -0.01443 -0.2592 1.076 1.35e-02
15  0.00215  0.003412 -0.06017 -0.01423  0.00568 -0.0678 1.588 9.64e-04
16  0.07301 -0.063407 -0.28162 -0.04035  0.08553 -0.3547 1.144 2.52e-02
17 -0.04812  0.074740 -0.13730  0.00412 -0.08866  0.2700 1.082 1.46e-02
18 -0.00159  0.003224 -0.00386 -0.00327 -0.00468  0.0108 1.363 2.42e-05
19  0.92507 -0.887507 -0.10576 -0.49081  0.04705  0.9997 1.111 1.89e-01
20  0.01345 -0.203406  0.27859  0.67796  0.37100 -0.9674 0.785 1.70e-01
21 -0.02581  0.009379 -0.10412  0.03393  0.17332 -0.2941 1.091 1.73e-02
22 -0.21910  0.251409  0.11613  0.03579 -0.30722  0.4307 1.076 3.66e-02
23 -0.35729  0.414710  0.28675 -0.08924 -0.37271  0.6511 1.704 8.59e-02
24 -0.20780  0.258913  0.16586 -0.48107  0.37720 -0.9842 0.721 1.73e-01
25  0.10514 -0.097553 -0.23133  0.15030 -0.24924  0.5297 1.207 5.58e-02
```

```
26 -0.02947  0.031264  0.00135  0.10701 -0.20053  0.2720 1.333 1.52e-02
27 -0.11711  0.076615  0.15152  0.10289  0.18923 -0.3154 1.406 2.04e-02
      hat inf
1  0.1722
2  0.1911
3  0.1519
4  0.3005   *
5  0.0983
6  0.1545
7  0.3423
8  0.3250   *
9  0.1281
10 0.1225
11 0.1994
12 0.2649
13 0.1587
14 0.0646
15 0.2085
16 0.1166
17 0.0698
18 0.0745
19 0.3113
20 0.2219
21 0.0806
22 0.1269
23 0.3731   *
24 0.2108
25 0.1976
26 0.1456
27 0.1886
```

```
> par(mfrow=c(2,2))
> plot(f1, ask=F)
```