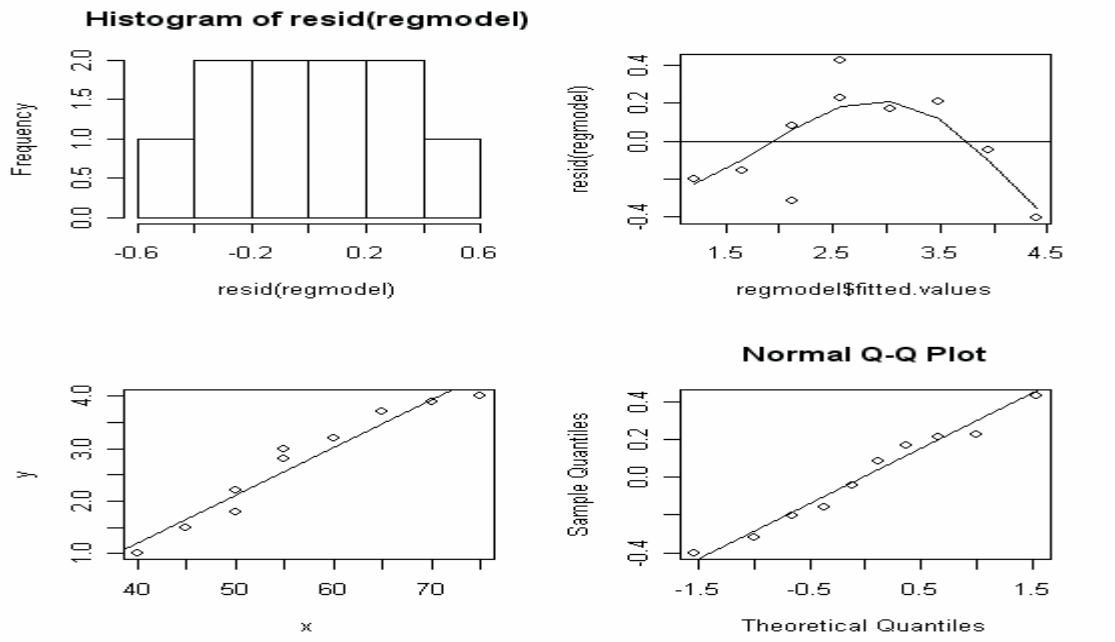
Linear Regression

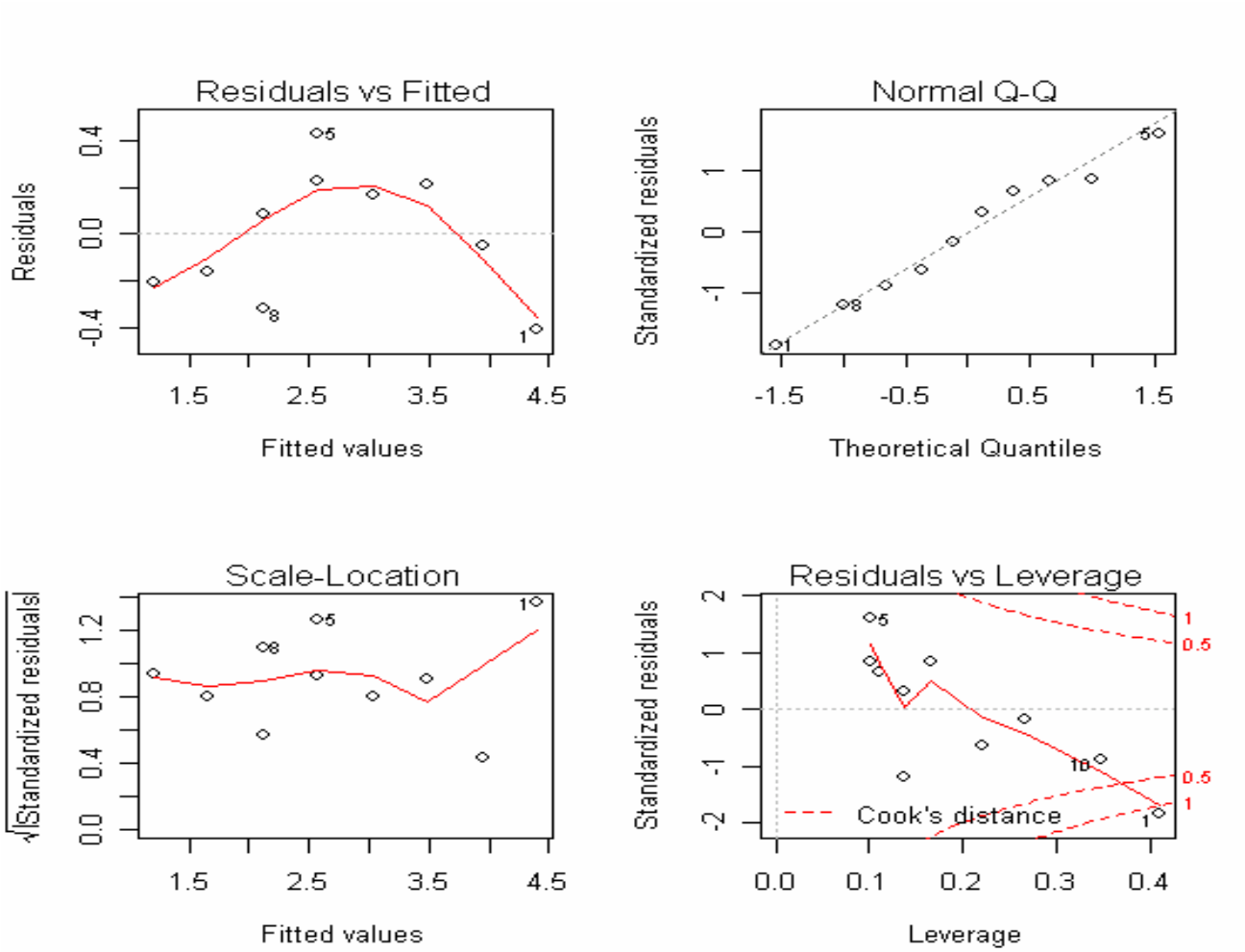
lm(formula, data, ...)
formula: a symbolic description of the model to be fit. The details of model specification are given below.

data: an optional data frame, list or a data frame.

Example:

```
x=c(75,70,65,60,55,55,50,50,45,40)
y=c(4,3.9,3.7,3.2,3,2.8,2.2,1.8,1.5,1)
regmodel=lm(y~x)
yy=data.frame(x,y)
pairs(yy)
regmodel=lm(y~x,yy)
par(mfrow=c(2,2))
hist(resid(regmodel))
> plot(regmodel$fitted.values,resid(regmodel))
> lines(lowess(regmodel$fitted.values,resid(regmodel)))
> abline(h=0)
> plot(x,y)
> abline(regmodel)
> qqnorm(resid(regmodel))
> qqline(resid(regmodel))
```





plot(regmodel)
summary(regmodel)

Call:
lm(formula = y ~ x)

Residuals:
```
     Min       1Q   Median       3Q      Max
-0.40227 -0.19002  0.01984  0.20181  0.42721
```

Coefficients:
```
             Estimate Std. Error t                    value Pr(>|t|)
(Intercept) -2.458277   0.491000  -5.007  0.00104 **
x            0.091474   0.008544  10.706 5.09e-06 ***
```

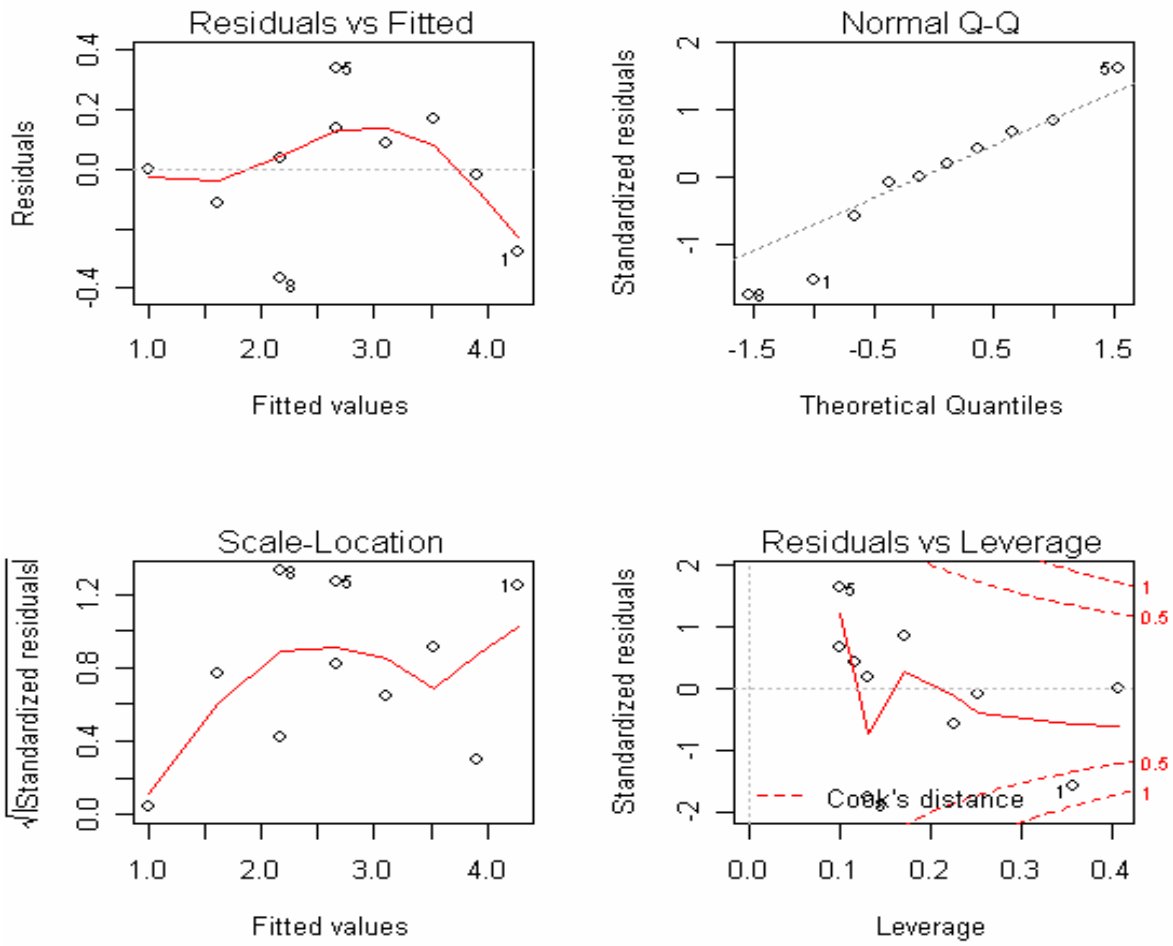
Residual standard error: 0.2837 on 8 degrees of freedom
Multiple R-Squared: 0.9348, Adjusted R-squared: 0.9266
F-statistic: 114.6 on 1 and 8 DF, p-value: 5.088e-06

cor(yy)
x 1.0000000 0.9668298
y 0.9668298 1.0000000
> sqrt(0.9348)
[1] 0.9668506

```
regmodel=lm(y~log(x))
plot(regmodel)
summary(regmodel)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-0.36339 -0.08996  0.01813  0.12647  0.33985

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -18.226      1.512  -12.06 2.07e-06 ***
log(x)         5.212      0.376   13.86 7.09e-07 ***
Residual standard error: 0.222 on 8 degrees of freedom
Multiple R-Squared:  0.96,      Adjusted R-squared: 0.955
F-statistic: 192.2 on 1 and 8 DF,  p-value: 7.09e-07
```



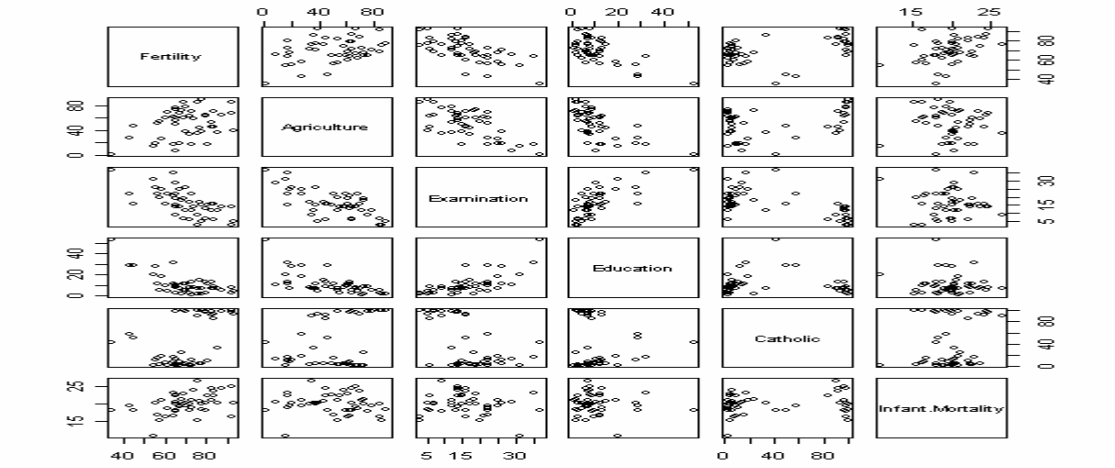
Multiple Regressions

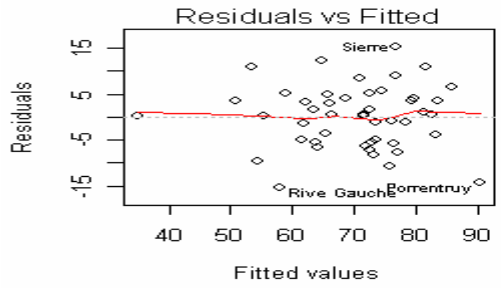
Example with swiss data

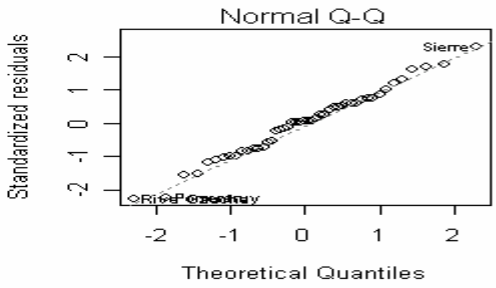
> pairs(swiss)

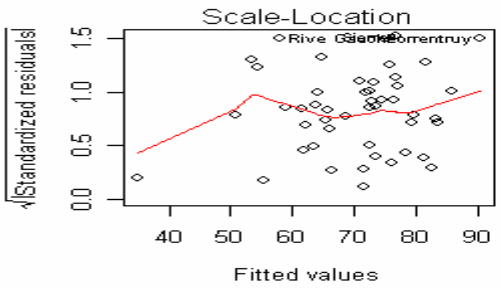
> cor(swiss)

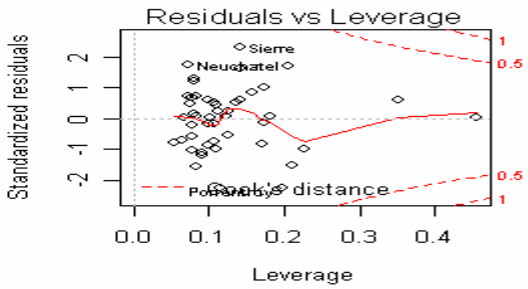
	Fertility
Fertility	**1.0000000**
Agriculture	**0.3530792**
Examination	**-0.6458827**
Education	**-0.6637889**
Catholic	**0.4636847**
Infant.Mortality	**0.4165560**











Multiple Regressions

summary(lm1 <- lm(Fertility ~ ., data = swiss))
summary(lm1 <- lm(Fertility ~ Agriculture+ Examination+ Education+ Catholic+ Infant.Mortality, data = swiss))

```
Call:
lm(formula = Fertility ~ ., data = swiss)
Residuals:
    Min      1Q  Median      3Q     Max
-15.2743  -5.2617   0.5032   4.1198  15.3213
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      66.91518   10.70604   6.250 1.91e-07 ***
Agriculture      -0.17211    0.07030  -2.448  0.01873 *
Examination      -0.25801    0.25388  -1.016  0.31546
Education        -0.87094    0.18303  -4.758 2.43e-05 ***
Catholic          0.10412    0.03526   2.953  0.00519 **
Infant.Mortality  1.07705    0.38172   2.822  0.00734 **
Residual standard error: 7.165 on 41 degrees of freedom
Multiple R-Squared: 0.7067,     Adjusted R-squared: 0.671
F-statistic: 19.76 on 5 and 41 DF,  p-value: 5.594e-10
```

AIC: the (generalized) Akaike Information Criterion for 'fit'.
AIC = ***-2*log-likelihood + k*npar*,**
where *npar* represents the number of parameters in the fitted model, and *k = 2* for the usual AIC, or *k = log(n)* (*n* the number of observations)
When comparing fitted objects, the smaller the AIC, the better the fit.

Single term additions

lm1 <- lm(Fertility ~ 1, data = swiss)
add1(lm1,~Agriculture+ Examination+ Education+ Catholic+ Infant.Mortality)

```
Model:
Fertility ~ 1
                 Df Sum of Sq    RSS    AIC
<none>                        7178.0  238.3
Agriculture       1     894.8 6283.1  234.1
Examination       1    2994.4 4183.6  215.0
Education         1    3162.7 4015.2  213.0
Catholic          1    1543.3 5634.7  229.0
Infant.Mortality  1    1245.5 5932.4  231.4
```

slm1 <- step(lm1)
Start: AIC= 190.69
Fertility ~ Agriculture + Examination + Education + Catholic +
Infant.Mortality

	Df	Sum of Sq	RSS	AIC
- Examination	1	53.0	2158.1	189.9
<none>			2105.0	190.7
- Agriculture	1	307.7	2412.8	195.1
- Infant.Mortality	1	408.8	2513.8	197.0
- Catholic	1	447.7	2552.8	197.8
- Education	1	1162.6	3267.6	209.4

Step: AIC= 189.86
Fertility ~ Agriculture + Education + Catholic + Infant.Mortality

	Df	Sum of Sq	RSS	AIC
<none>			2158.1	189.9
- Agriculture	1	264.2	2422.2	193.3
- Infant.Mortality	1	409.8	2567.9	196.0
- Catholic	1	956.6	3114.6	205.1
- Education	1	2250.0	4408.0	221.4

Single term deletions
drop1(lm1)
Model:
Fertility ~ Agriculture + Examination + Education + Catholic +
Infant.Mortality

	Df	Sum of Sq	RSS	AIC
<none>			2105.0	190.7
Agriculture	1	307.7	2412.8	195.1
Examination	1	53.0	2158.1	189.9
Education	1	1162.6	3267.6	209.4
Catholic	1	447.7	2552.8	197.8
Infant.Mortality	1	408.8	2513.8	197.0

```
lm4 <- lm(Fertility ~ Agriculture+ Education+ Catholic+
Infant.Mortality, data = swiss)
> lm1 <- lm(Fertility ~ Agriculture+ Examination+ Education+
Catholic+ Infant.Mortality, data = swiss)
> anova(lm1,lm4)
Analysis of Variance Table

Model 1: Fertility ~ Agriculture + Examination + Education + Catholic +
    Infant.Mortality
Model 2: Fertility ~ Agriculture + Education + Catholic +
Infant.Mortality
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     41 2105.04
2     42 2158.07 -1    -53.03 1.0328 0.3155

> lm3 <- lm(Fertility ~  Education+ Catholic+ Infant.Mortality, data
= swiss)
> anova(lm1,lm3)
Analysis of Variance Table

Model 1: Fertility ~ Agriculture + Examination + Education + Catholic +
    Infant.Mortality
Model 2: Fertility ~ Education + Catholic + Infant.Mortality
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     41 2105.0
2     43 2422.3 -2    -317.2 3.0891 0.05628 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```