Simple linear regression model

The basic model for linear regression is that pairs of data, (x_i , y_i), are related through the equation

$$\boldsymbol{y_i = \beta_o + \beta_1 x_i + \epsilon_i}$$

where β_0 and β_1 are unknown constants and can be estimated from the data.

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)

y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)

plot(x,y)                          # make a plot

abline(lm(y ~ x))                  # plot the regression line
```

or

```
fit<-simple.lm(x,y)

summary(fit)                       #to obtain summary of the model
```

Testing the assumptions of the model

We have to check the validity of the model before accepting it for the intended purpose. Visual inspection of residual is the most popular way to test the validly of a regression model. The assumption on the errors being independent and identical distributed and also normally distributed. The variance of residuals are not uniform (constant), but they should show no serial correlations.

We can test for normality with *histograms, boxplots* and *normal probability plots*. We can test for correlations by looking if there are trends in the data. This can be done with plots of the *residuals vs. time and order*. We can test the assumption that the errors have the same variance with *plots of residuals vs. time order and fitted values*.

The plot command will do these tests for us if we give it the result of the regression

par(mfrow=c(2,2) *#place 4 on one graph*

plot(fit)

Residuals vs. fitted

This plots the fitted (b y) values against the residuals. Look for spread around the line y = 0 and no obvious trend.

Normal qqplot

The residuals are normal if this graph falls close to a straight line where straight line indicates the normally distributed data.

Scale-Location

This plot shows the square root of the standardized residuals. The tallest points are the largest residuals.

Cook's distance

This plot identifies points which have a lot of influence in the regression line.