

Regression Analysis lab 7

1 Indicator variables

1.1 Import data

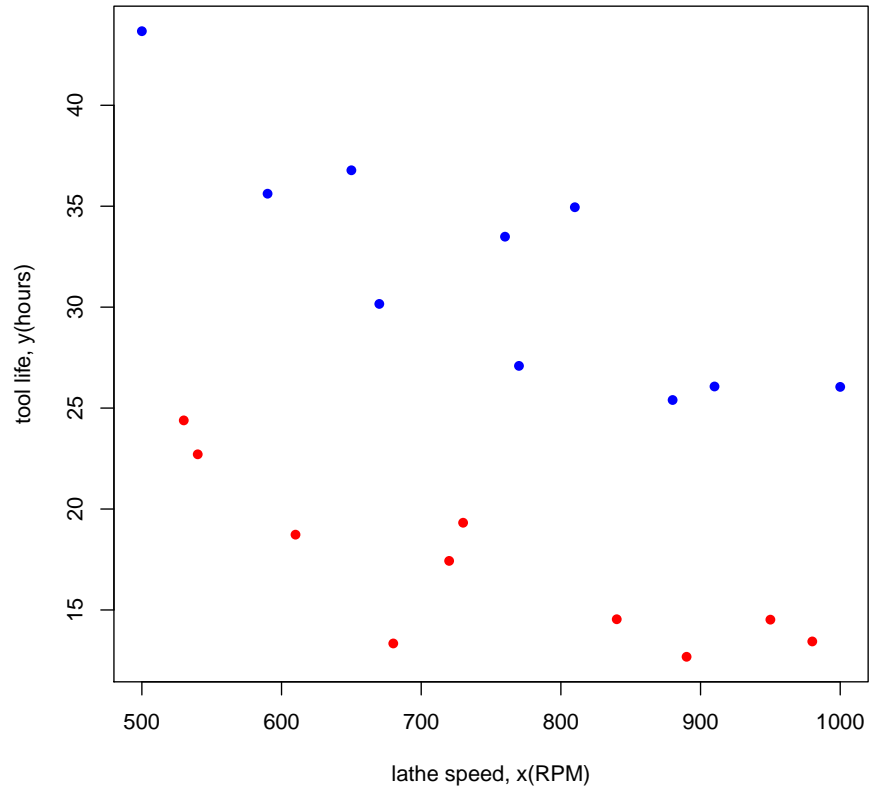
```
tool<-read.csv(file="D:/chilo/Regression 7/tool.csv", header=T)
tool
```

	life	speed	type
1	18.73	610	A
2	14.52	950	A
3	17.43	720	A
4	14.54	840	A
5	13.44	980	A
6	24.39	530	A
7	13.34	680	A
8	22.71	540	A
9	12.68	890	A
10	19.32	730	A
11	30.16	670	B
12	27.09	770	B
13	25.40	880	B
14	26.05	1000	B
15	33.49	760	B
16	35.62	590	B
17	26.07	910	B
18	36.78	650	B
19	34.95	810	B
20	43.67	500	B

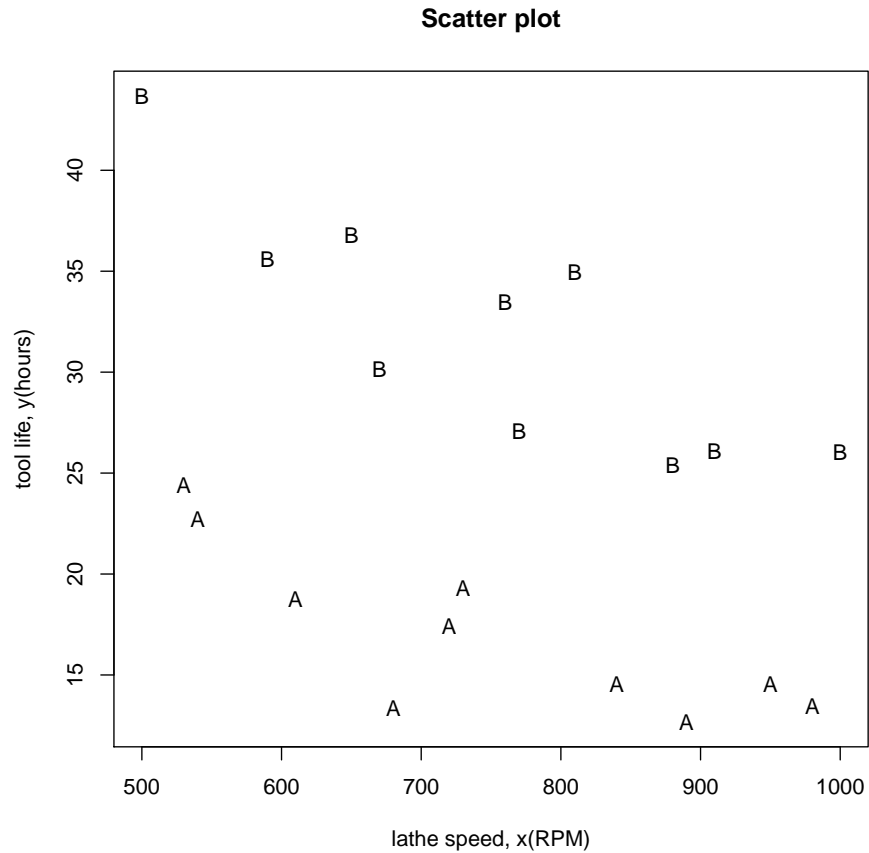
2 scatter plot by group

```
attach(tool)
plot(speed, life, pch=16, col=c("red", "blue")[type],
      xlab="lathe speed, x(RPM)", ylab="tool life, y(hours)",
      main="Scatter plot")
```

Scatter plot



```
plot(speed, life, pch=16, type="n",  
      xlab="lathe speed, x(RPM)", ylab="tool life, y(hours)",  
      main="Scatter plot")  
text(speed, life, type)
```



3 model 1 with y and x1

```
attach(tool)
```

The following objects are masked from tool (position 3):

```
life, speed, type
```

```
tfit1 <- lm(life ~ speed, data=tool)
```

```
summary(tfit1)
```

Call:

```
lm(formula = life ~ speed, data = tool)
```

```

Residuals:
  Min       1Q   Median       3Q      Max
-12.973  -7.300  -0.928   7.233  12.777

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  43.6167    9.6032    4.54  0.00025 ***
speed       -0.0254    0.0125   -2.03  0.05760 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.44 on 18 degrees of freedom
Multiple R-squared:  0.186, Adjusted R-squared:  0.141
F-statistic: 4.11 on 1 and 18 DF, p-value: 0.0576

tfit1$fit # fitted values
  1    2    3    4    5    6    7    8    9   10   11   12
28.09 19.44 25.30 22.24 18.68 30.13 26.31 29.88 20.97 25.04 26.57 24.02
 13   14   15   16   17   18   19   20
21.22 18.17 24.28 28.60 20.46 27.08 23.00 30.89

tfit1$res # residuals
  1    2    3    4    5    6    7    8    9
-9.364 -4.922 -7.865 -7.702 -5.239 -5.740 -12.973 -7.166 -8.289
 10   11   12   13   14   15   16   17   18
-5.721  3.593  3.067  4.176  7.880  9.213  7.017  5.610  9.704
 19   20
11.945 12.777

summary(tfit1)

Call:
lm(formula = life ~ speed, data = tool)

Residuals:
  Min       1Q   Median       3Q      Max
-12.973  -7.300  -0.928   7.233  12.777

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  43.6167    9.6032    4.54  0.00025 ***
speed       -0.0254    0.0125   -2.03  0.05760 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 8.44 on 18 degrees of freedom
Multiple R-squared: 0.186, Adjusted R-squared: 0.141
F-statistic: 4.11 on 1 and 18 DF, p-value: 0.0576

anova(tfit1)

Analysis of Variance Table

Response: life
      Df Sum Sq Mean Sq F value Pr(>F)
speed   1    293   293.0    4.11 0.058 .
Residuals 18  1282    71.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

confint(tfit1, level=0.95)

              2.5 %    97.5 %
(Intercept) 23.44108 6.379e+01
speed       -0.05181 9.121e-04

R2_1<-summary(tfit1)$r.squared
R2_1

[1] 0.186

R2.adj_1<-summary(tfit1)$adj.r.squared
R2.adj_1

[1] 0.1408

sigmahat_1<-summary(tfit1)$sigma
sigmahat_1

[1] 8.44

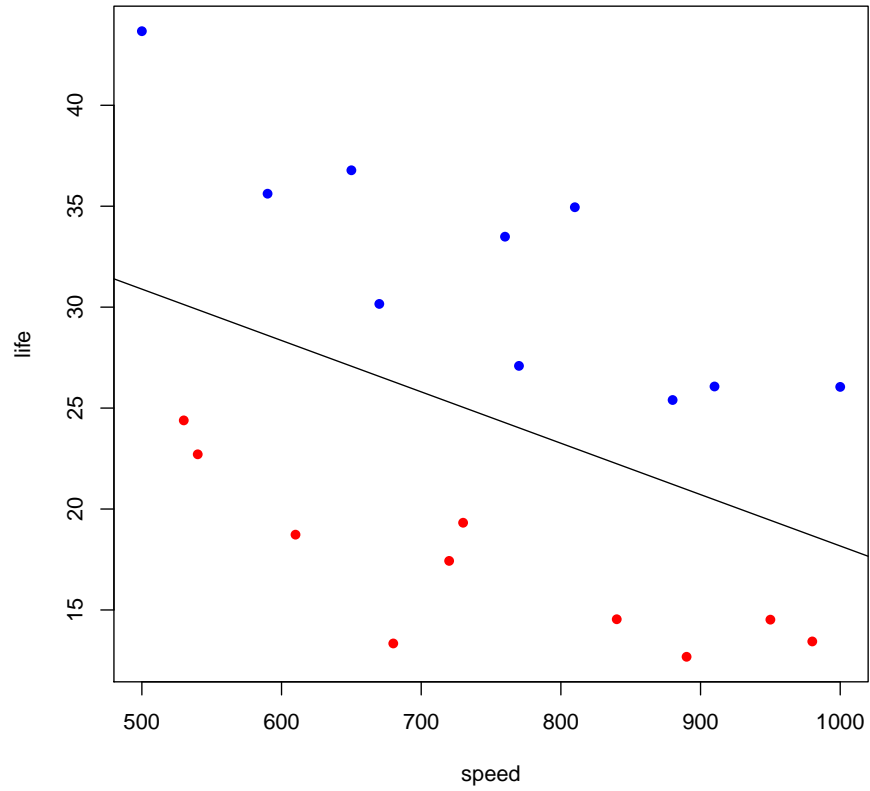
sigmahat2_1<-sigmahat_1^2
MSE_1<-sigmahat2_1
MSE_1

[1] 71.23

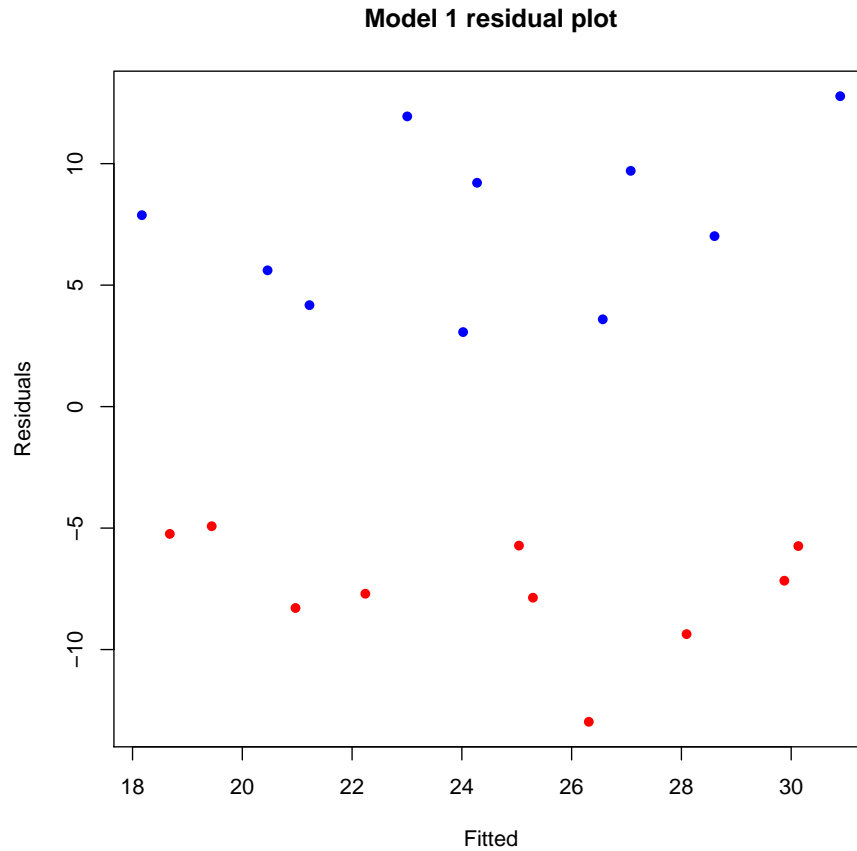
plot(life ~ speed, pch=16, col=c("red", "blue")[type],
      main="Scatter plot")
abline(reg=tfit1)

```

Scatter plot



```
plot(tfit1$fit,tfit1$res, pch=16, col=c("red","blue")[type],  
      xlab="Fitted",ylab="Residuals",  
      main="Model 1 residual plot")
```



4 model 2 with y and x1, x2

```
attach(tool)
```

```
The following objects are masked from tool (position 3):
```

```
life, speed, type
```

```
The following objects are masked from tool (position 4):
```

```
life, speed, type
```

```
tfit2 <- lm(life ~ speed + type, data=tool)
```

```
summary(tfit2)
```

```

Call:
lm(formula = life ~ speed + type, data = tool)

Residuals:
    Min       1Q   Median       3Q      Max
-5.553 -1.787 -0.002  1.839  4.984

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 36.98560    3.51038   10.54 7.2e-09 ***
speed       -0.02661    0.00452   -5.89 1.8e-05 ***
typeB       15.00425    1.35967   11.04 3.6e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.04 on 17 degrees of freedom
Multiple R-squared:  0.9, Adjusted R-squared:  0.889
F-statistic: 76.7 on 2 and 17 DF, p-value: 3.09e-09

tfit2$fit # fitted values
  1    2    3    4    5    6    7    8    9   10   11   12
20.76 11.71 17.83 14.64 10.91 22.88 18.89 22.62 13.31 17.56 34.16 31.50
 13   14   15   16   17   18   19   20
28.58 25.38 31.77 36.29 27.78 34.70 30.44 38.69

tfit2$res # residuals
  1    2    3    4    5    6    7    8
-2.02519 2.81127 -0.39840 -0.09553 2.52948 1.50623 -5.55268 0.09230
  9   10   11   12   13   14   15   16
-0.62517 1.75768 -4.00301 -4.41228 -3.17549 0.66738 1.72164 -0.67159
 17   18   19   20
-1.70727 2.08485 4.51200 4.98376

summary(tfit2)

Call:
lm(formula = life ~ speed + type, data = tool)

Residuals:
    Min       1Q   Median       3Q      Max
-5.553 -1.787 -0.002  1.839  4.984

Coefficients:
            Estimate Std. Error t value Pr(>|t|)

```



```

(Intercept) 36.98560    3.51038    10.54  7.2e-09 ***
speed       -0.02661    0.00452    -5.89  1.8e-05 ***
typeB       15.00425    1.35967    11.04  3.6e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.04 on 17 degrees of freedom
Multiple R-squared:  0.9, Adjusted R-squared:  0.889
F-statistic: 76.7 on 2 and 17 DF,  p-value: 3.09e-09

anova(tfit2)

Analysis of Variance Table

Response: life
      Df Sum Sq Mean Sq F value Pr(>F)
speed   1   293     293    31.7 3.0e-05 ***
type    1  1125    1125   121.8 3.6e-09 ***
Residuals 17   157         9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

confint(tfit2, level=0.95)

           2.5 %   97.5 %
(Intercept) 29.57934 44.39186
speed       -0.03614 -0.01707
typeB       12.13560 17.87290

R2_2<-summary(tfit2)$r.squared
R2_2

[1] 0.9003

R2.adj_2<-summary(tfit2)$adj.r.squared
R2.adj_2

[1] 0.8886

sigmahat_2<-summary(tfit2)$sigma
sigmahat_2

[1] 3.039

sigmahat2_2<-sigmahat_2^2
MSE_2<-sigmahat2_2
MSE_2

```

```

[1] 9.239

beta.0<-summary(tfit2)$coef[1,1]
beta.1<-summary(tfit2)$coef[2,1]
beta.2<-summary(tfit2)$coef[3,1]
beta.0

[1] 36.99

beta.1

[1] -0.02661

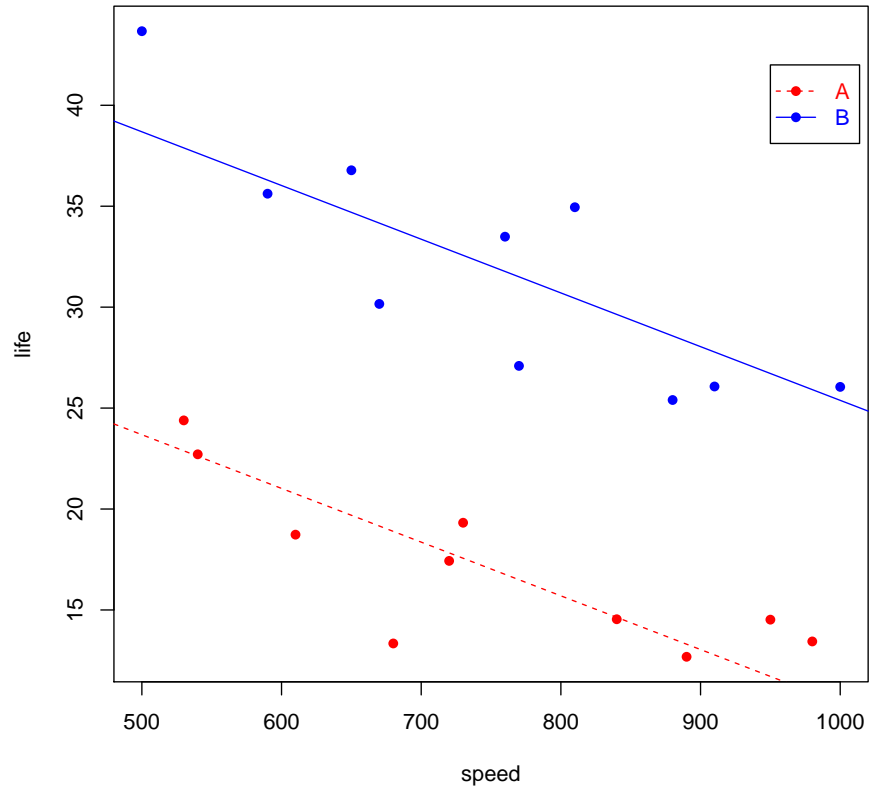
beta.2

[1] 15

plot(life ~ speed, pch = 16,, col=c("red","blue")[type],
      main="Scatter plot")
abline(beta.0, beta.1, lty=2, col="red")
abline(beta.0+beta.2, beta.1, lty=1, col="blue")
legend(950, 42, c("A", "B"), col = c("red", "blue"),
       text.col = c("red", "blue"), lty = c(2, 1),
       pch = c(16, 16), merge = TRUE)

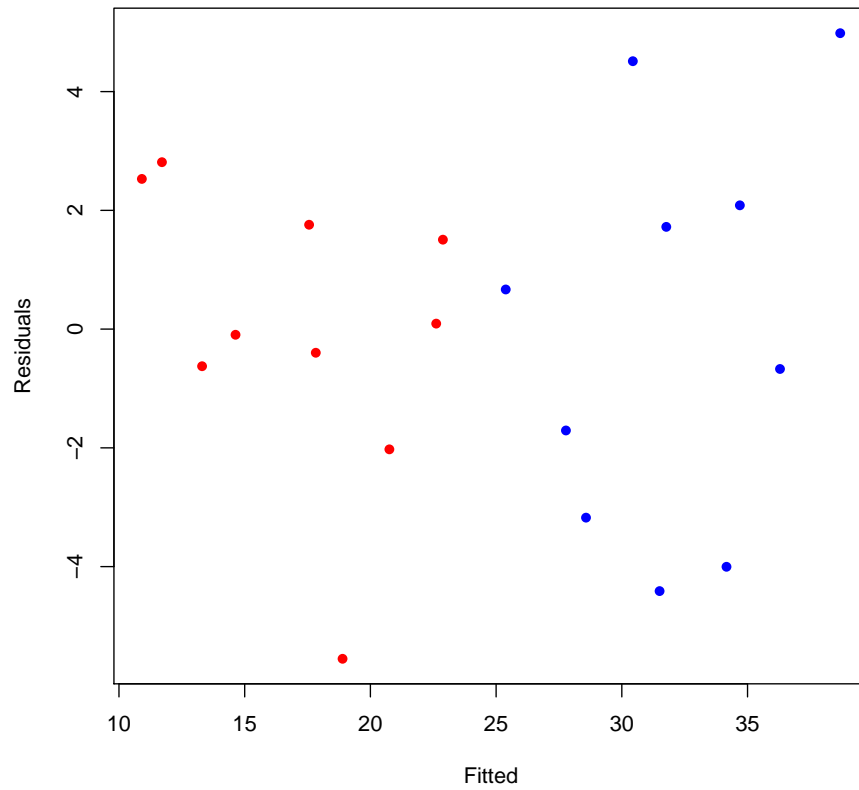
```

Scatter plot



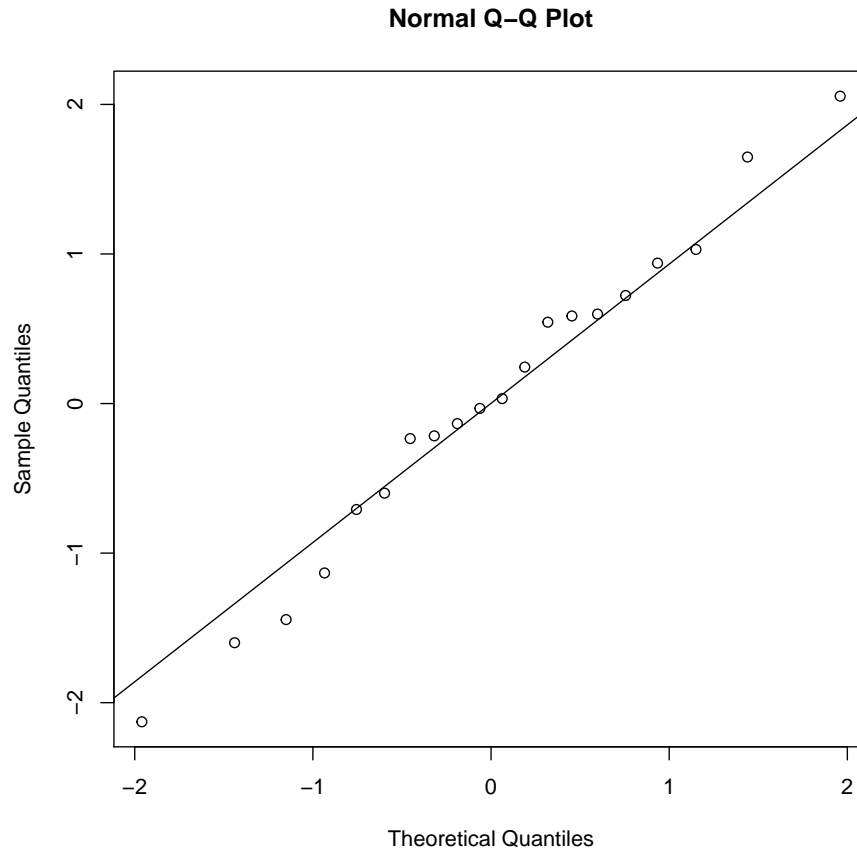
```
plot(tfit2$fit,tfit2$res, pch=16, col=c("red","blue")[type],
     xlab="Fitted",ylab="Residuals",
     main="Model 2 residual plot")
```

Model 2 residual plot



```
t1<-rstudent(tfit2)
t1
      1      2      3      4      5      6      7      8
-0.70850  1.03025 -0.13423 -0.03249  0.93903  0.54385 -2.12781  0.03283
      9     10     11     12     13     14     15     16
-0.21614  0.59815 -1.44449 -1.59924 -1.13270  0.24381  0.58543 -0.23421
     17     18     19     20
-0.59899  0.72245  1.64823  2.05498

qqnorm(t1)
qqline(t1)
```



5 model 3 with y and x1, x2, x1x2

```
attach(tool)
```

The following objects are masked from tool (position 3):

```
life, speed, type
```

The following objects are masked from tool (position 4):

```
life, speed, type
```

The following objects are masked from tool (position 5):

```
life, speed, type
```

```
tfit3 <- lm(life ~ speed + type + speed*type, data=tool)
```

```
summary(tfit3)

Call:
lm(formula = life ~ speed + type + speed * type, data = tool)

Residuals:
    Min       1Q   Median       3Q      Max
-5.175 -1.500  0.485  1.783  4.865

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.77476    4.63347   7.07 2.6e-06 ***
speed       -0.02097    0.00607  -3.45 0.0033 **
typeB       23.97059    6.76897   3.54 0.0027 **
speed:typeB -0.01194    0.00884  -1.35 0.1955
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.97 on 16 degrees of freedom
Multiple R-squared:  0.91, Adjusted R-squared:  0.894
F-statistic: 54.3 on 3 and 16 DF, p-value: 1.32e-08

tfit3$fit # fitted values
   1    2    3    4    5    6    7    8    9   10   11   12
19.98 12.85 17.68 15.16 12.22 21.66 18.52 21.45 14.11 17.47 34.69 31.40
   13   14   15   16   17   18   19   20
27.78 23.83 31.73 37.33 26.79 35.35 30.08 40.29

tfit3$res # residuals
   1    2    3    4    5    6    7    8    9
-1.2529 1.6670 -0.2462 -0.6198 1.2161 2.7295 -5.1750 1.2592 -1.4313
   10   11   12   13   14   15   16   17   18
1.8535 -4.5328 -4.3114 -2.3808 2.2189 1.7595 -1.7059 -0.7234 1.4289
   19   20
4.8652 3.3818

summary(tfit3)

Call:
lm(formula = life ~ speed + type + speed * type, data = tool)

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

```

-5.175 -1.500 0.485 1.783 4.865

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.77476    4.63347    7.07 2.6e-06 ***
speed       -0.02097    0.00607   -3.45 0.0033 **
typeB       23.97059    6.76897    3.54 0.0027 **
speed:typeB -0.01194    0.00884   -1.35 0.1955
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.97 on 16 degrees of freedom
Multiple R-squared:  0.91, Adjusted R-squared:  0.894
F-statistic: 54.3 on 3 and 16 DF,  p-value: 1.32e-08

anova(tfit3)

Analysis of Variance Table

Response: life
          Df Sum Sq Mean Sq F value Pr(>F)
speed     1    293     293    33.25 2.9e-05 ***
type      1   1125    1125   127.68 4.9e-09 ***
speed:type 1     16      16     1.82   0.2
Residuals 16    141      9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

confint(tfit3, level=0.95)

              2.5 %    97.5 %
(Intercept) 22.95224 42.597281
speed       -0.03385 -0.008094
typeB       9.62101 38.320175
speed:typeB -0.03069 0.006800

R2_3<-summary(tfit3)$r.squared
R2_3

[1] 0.9105

R2.adj_3<-summary(tfit3)$adj.r.squared
R2.adj_3

[1] 0.8937

sigmahat_3<-summary(tfit3)$sigma
sigmahat_3

```

```

[1] 2.968

sigmahat2_3<-sigmahat_3^2
MSE_3<-sigmahat2_3
MSE_3

[1] 8.811

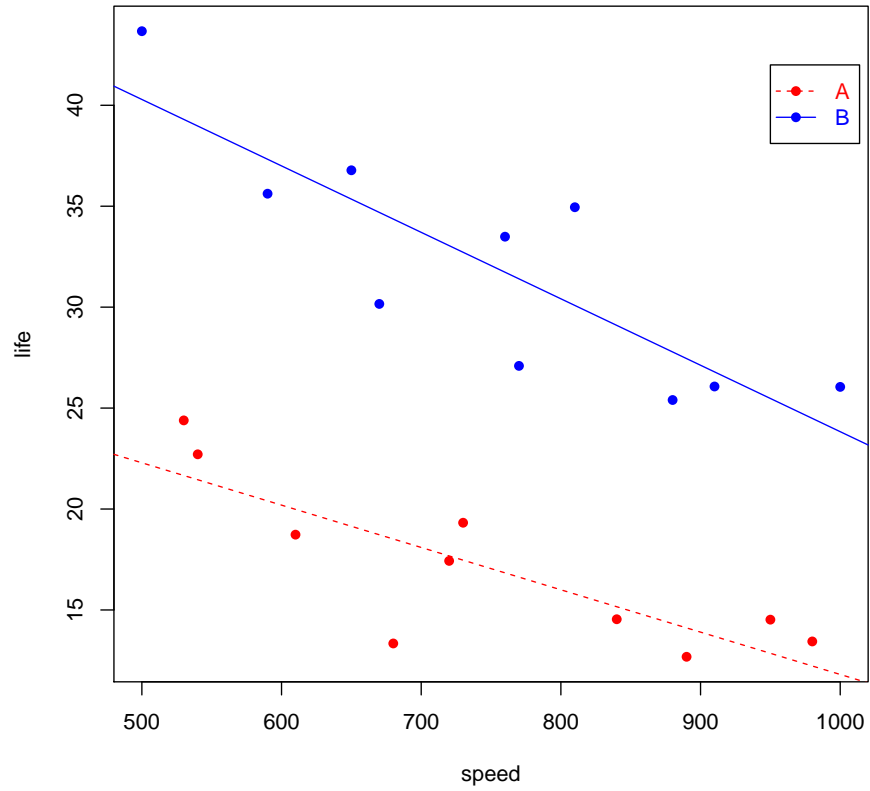
toolA<-tool[type=="A",]
toolA
  life speed type
1  18.73   610   A
2  14.52   950   A
3  17.43   720   A
4  14.54   840   A
5  13.44   980   A
6  24.39   530   A
7  13.34   680   A
8  22.71   540   A
9  12.68   890   A
10 19.32   730   A

toolB<-tool[type=="B",]
toolB
  life speed type
11 30.16   670   B
12 27.09   770   B
13 25.40   880   B
14 26.05  1000   B
15 33.49   760   B
16 35.62   590   B
17 26.07   910   B
18 36.78   650   B
19 34.95   810   B
20 43.67   500   B

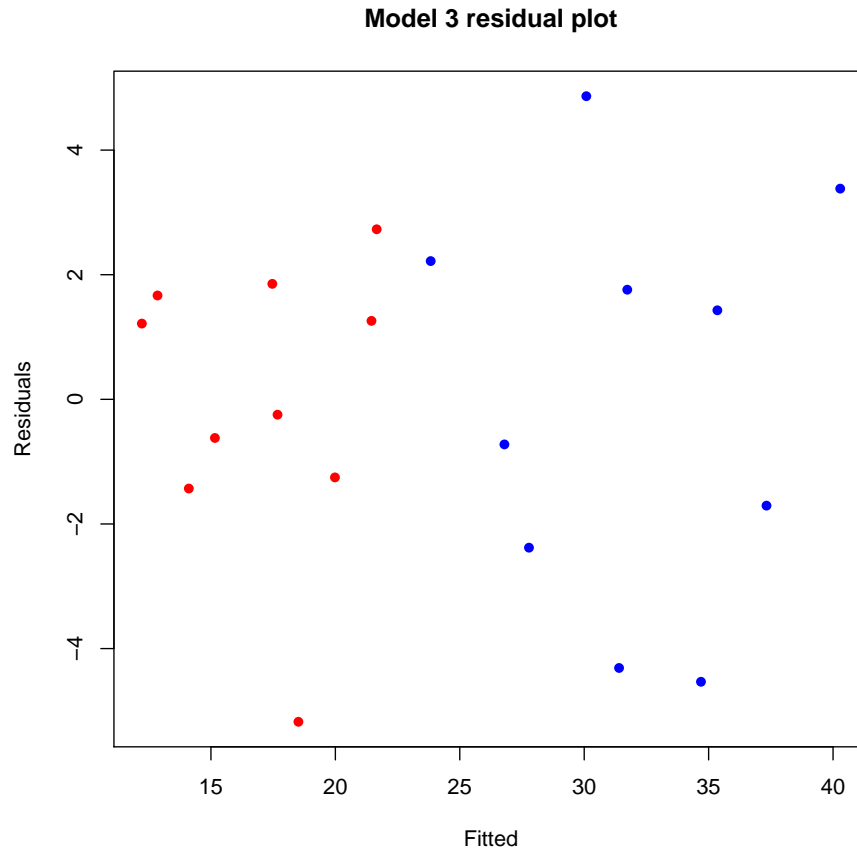
plot(life ~ speed, pch = 16,, col=c("red","blue")[type],
     main="Scatter plot")
abline(lm(toolA$life ~ toolA$speed), lty=2, col="red")
abline(lm(toolB$life ~ toolB$speed), lty=1, col="blue")
legend(950, 42, c("A", "B"), col = c("red", "blue"),
      text.col = c("red", "blue"), lty = c(2, 1), pch = c(16, 16),
      merge = TRUE)

```


Scatter plot



```
plot(tfit3$fit,tfit3$res, pch=16, col=c("red","blue")[type],  
      xlab="Fitted",ylab="Residuals",  
      main="Model 3 residual plot")
```



6 partial F tests

6.1 test whether two regression lines are identical

```
attach(tool)
```

```
The following objects are masked from tool (position 3):
```

```
life, speed, type
```

```
The following objects are masked from tool (position 4):
```

```
life, speed, type
```

```
The following objects are masked from tool (position 5):
```

```
life, speed, type
```

```

The following objects are masked from tool (position 6):

  life, speed, type

tfit1 <- lm(life ~ speed, data=tool) # Reduced model
tfit3 <- lm(life ~ speed + type + speed*type, data=tool) # Full model
anova(tfit1,tfit3)

Analysis of Variance Table

Model 1: life ~ speed
Model 2: life ~ speed + type + speed * type
  Res.Df  RSS Df Sum of Sq   F Pr(>F)
1      18 1282
2      16  141  2     1141 64.8 2.1e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

6.2 test whether the slopes of two regression lines are equal

```

attach(tool)

The following objects are masked from tool (position 3):

  life, speed, type
The following objects are masked from tool (position 4):

  life, speed, type
The following objects are masked from tool (position 5):

  life, speed, type
The following objects are masked from tool (position 6):

  life, speed, type
The following objects are masked from tool (position 7):

  life, speed, type

tfit2 <- lm(life ~ speed + type, data=tool) # Reduced model
tfit3 <- lm(life ~ speed + type + speed*type, data=tool) # Full model
anova(tfit2,tfit3)

Analysis of Variance Table

```

```
Model 1: life ~ speed + type
Model 2: life ~ speed + type + speed * type
  Res.Df RSS Df Sum of Sq    F Pr(>F)
1     17 157
2     16 141  1     16.1 1.82  0.2
```