Regression Analysis lab 3

1 Multiple linear regression

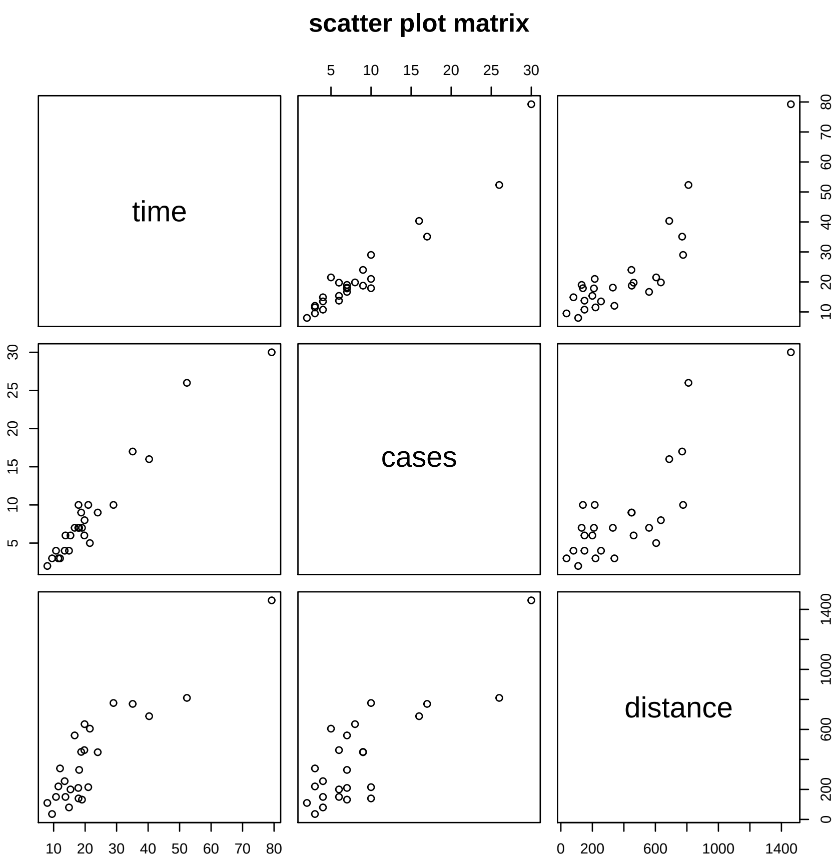
1.1 Import data

```
delivery<-read.csv(file="D:/chilo/Regression 3/delivery.csv", header=T)
delivery
```

```
   observation  time cases distance
1            1 16.68     7      560
2            2 11.50     3      220
3            3 12.03     3      340
4            4 14.88     4       80
5            5 13.75     6      150
6            6 18.11     7      330
7            7  8.00     2      110
8            8 17.83     7      210
9            9 79.24    30     1460
10          10 21.50     5      605
11          11 40.33    16      688
12          12 21.00    10      215
13          13 13.50     4      255
14          14 19.75     6      462
15          15 24.00     9      448
16          16 29.00    10      776
17          17 15.35     6      200
18          18 19.00     7      132
19          19  9.50     3       36
20          20 35.10    17      770
21          21 17.90    10      140
22          22 52.32    26      810
23          23 18.75     9      450
24          24 19.83     8      635
25          25 10.75     4      150
```

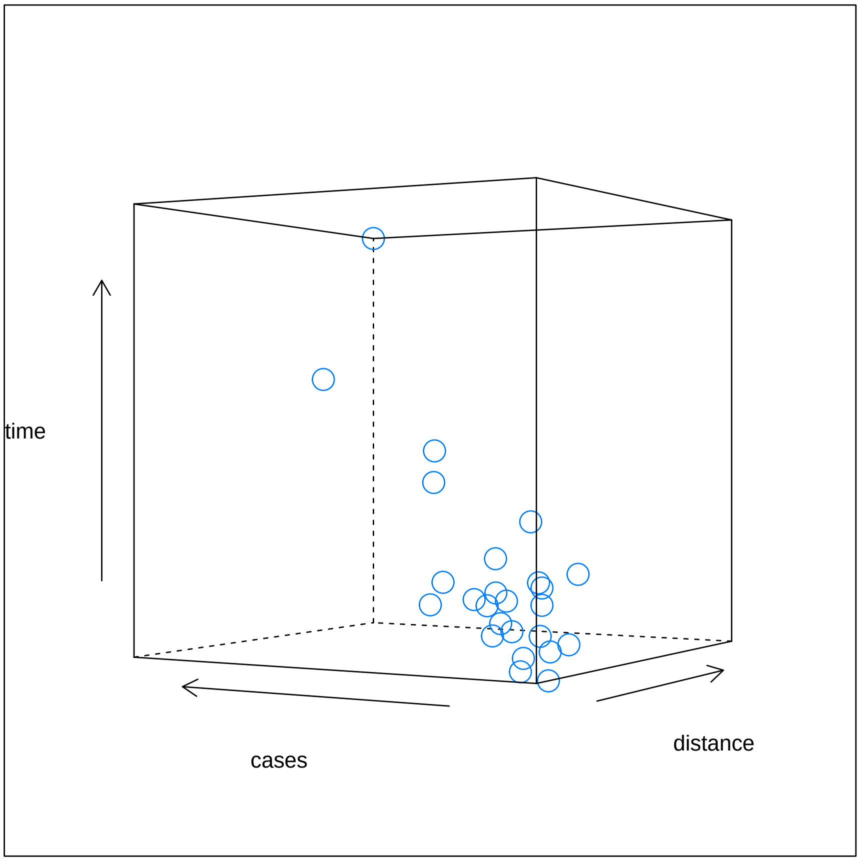
1.2 Scatterplot matrix

```
plot(delivery[,-1], main=("scatter plot matrix"))
```



1.3 3-D Scatter plot

```
library(lattice)
cloud(time ~ distance * cases, data = delivery, pch=1, cex=2,
      screen = list(z = 120, x = -90, y = -60))
```



1.4 Compute estimated coefficients, fitted values and residuals by formula

```
attach(delivery)
n<-length(delivery$time)
n

[1] 25

delivery[,-c(1,2)]

   cases distance
1      7      560
2      3      220
3      3      340
```

```
4      4      80
5      6     150
6      7     330
7      2     110
8      7     210
9     30    1460
10     5     605
11    16     688
12    10     215
13     4     255
14     6     462
15     9     448
16    10     776
17     6     200
18     7     132
19     3      36
20    17     770
21    10     140
22    26     810
23     9     450
24     8     635
25     4     150
```

```
X<-cbind(1,delivery[,-c(1,2)])
X
```

```
   1 cases distance
1  1     7      560
2  1     3      220
3  1     3      340
4  1     4       80
5  1     6      150
6  1     7      330
7  1     2      110
8  1     7      210
9  1    30     1460
10 1     5      605
11 1    16      688
12 1    10      215
13 1     4      255
14 1     6      462
15 1     9      448
16 1    10      776
17 1     6      200
18 1     7      132
19 1     3       36
```

```
20 1    17      770
21 1    10      140
22 1    26      810
23 1     9      450
24 1     8      635
25 1     4      150
```

```
y<-delivery$time
y
```

```
 [1] 16.68 11.50 12.03 14.88 13.75 18.11  8.00 17.83 79.24 21.50 40.33
[12] 21.00 13.50 19.75 24.00 29.00 15.35 19.00  9.50 35.10 17.90 52.32
[23] 18.75 19.83 10.75
```

```
X <- as.matrix(X)
t(X) %*% X
```

```
             1  cases distance
1           25    219    10232
cases      219   3055   133899
distance 10232 133899  6725688
```

```
XtXi <- solve(t(X) %*% X)
XtXi
```

```
                  1      cases   distance
1         1.132e-01 -4.449e-03 -8.367e-05
cases    -4.449e-03  2.744e-03 -4.786e-05
distance -8.367e-05 -4.786e-05  1.229e-06
```

```
Xty<-t(X) %*% y
Xty
```

```
             [,1]
1           559.6
cases      7375.4
distance 337071.7
```

```
beta<-XtXi %*% t(X) %*% y
beta
```

```
            [,1]
1        2.34123
cases    1.61591
distance 0.01438
```

```
yhat<-X %*% beta
yhat

        [,1]
 [1,] 21.708
 [2,] 10.354
 [3,] 12.080
 [4,]  9.956
 [5,] 14.194
 [6,] 18.400
 [7,]  7.155
 [8,] 16.673
 [9,] 71.820
[10,] 19.124
[11,] 38.093
[12,] 21.593
[13,] 12.473
[14,] 18.682
[15,] 23.329
[16,] 29.663
[17,] 14.914
[18,] 15.551
[19,]  7.707
[20,] 40.888
[21,] 20.514
[22,] 56.007
[23,] 23.358
[24,] 24.403
[25,] 10.963
```

```
e<-y-yhat
e

          [,1]
 [1,] -5.02808
 [2,]  1.14639
 [3,] -0.04979
 [4,]  4.92435
 [5,] -0.44440
 [6,] -0.28957
 [7,]  0.84462
 [8,]  1.15660
 [9,]  7.41971
[10,]  2.37641
[11,]  2.23749
```

```
[12,] -0.59304
[13,]  1.02701
[14,]  1.06754
[15,]  0.67120
[16,] -0.66293
[17,]  0.43636
[18,]  3.44862
[19,]  1.79319
[20,] -5.78797
[21,] -2.61418
[22,] -3.68653
[23,] -4.60757
[24,] -4.57285
[25,] -0.21258
```

```
cbind(y,yhat,e)
```

```
          y
 [1,] 16.68 21.708 -5.02808
 [2,] 11.50 10.354  1.14639
 [3,] 12.03 12.080 -0.04979
 [4,] 14.88  9.956  4.92435
 [5,] 13.75 14.194 -0.44440
 [6,] 18.11 18.400 -0.28957
 [7,]  8.00  7.155  0.84462
 [8,] 17.83 16.673  1.15660
 [9,] 79.24 71.820  7.41971
[10,] 21.50 19.124  2.37641
[11,] 40.33 38.093  2.23749
[12,] 21.00 21.593 -0.59304
[13,] 13.50 12.473  1.02701
[14,] 19.75 18.682  1.06754
[15,] 24.00 23.329  0.67120
[16,] 29.00 29.663 -0.66293
[17,] 15.35 14.914  0.43636
[18,] 19.00 15.551  3.44862
[19,]  9.50  7.707  1.79319
[20,] 35.10 40.888 -5.78797
[21,] 17.90 20.514 -2.61418
[22,] 52.32 56.007 -3.68653
[23,] 18.75 23.358 -4.60757
[24,] 19.83 24.403 -4.57285
[25,] 10.75 10.963 -0.21258
```

1.5 Compute the estimate of variance

```
yty<-t(y) %*% y
yty

      [,1]
[1,] 18311

t(beta) %*% t(X) %*% y

      [,1]
[1,] 18077

SSE<-yty-t(beta) %*% t(X) %*% y
SSE

      [,1]
[1,] 233.7

sigmahat2<-SSE/(n-3)
sigmahat2

      [,1]
[1,] 10.62

sigmahat2<-sum(e^2)/(n-3)
sigmahat2

[1] 10.62
```

1.6 Fit a multiple linear regression

```
attach(delivery)

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

    cases, distance, observation, time

dfit <- lm(time ~ cases + distance, data=delivery)
summary(dfit)


Call:
lm(formula = time ~ cases + distance, data = delivery)
```

```
Residuals:
   Min     1Q Median     3Q    Max
-5.788 -0.663  0.436  1.157  7.420

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.34123    1.09673    2.13  0.04417 *
cases        1.61591    0.17073    9.46  3.3e-09 ***
distance     0.01438    0.00361    3.98  0.00063 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.26 on 22 degrees of freedom
Multiple R-squared:  0.96,Adjusted R-squared:  0.956
F-statistic:  261 on 2 and 22 DF,  p-value: 4.69e-16
```

```
dfit$fit
```

```
     1      2      3      4      5      6      7      8      9     10
21.708 10.354 12.080  9.956 14.194 18.400  7.155 16.673 71.820 19.124
    11     12     13     14     15     16     17     18     19     20
38.093 21.593 12.473 18.682 23.329 29.663 14.914 15.551  7.707 40.888
    21     22     23     24     25
20.514 56.007 23.358 24.403 10.963
```

```
dfit$res
```

```
       1        2        3        4        5        6        7        8
-5.02808  1.14639 -0.04979  4.92435 -0.44440 -0.28957  0.84462  1.15660
       9       10       11       12       13       14       15       16
 7.41971  2.37641  2.23749 -0.59304  1.02701  1.06754  0.67120 -0.66293
      17       18       19       20       21       22       23       24
 0.43636  3.44862  1.79319 -5.78797 -2.61418 -3.68653 -4.60757 -4.57285
      25
-0.21258
```

```
sigmahat2<-sum(dfit$res^2)/(n-3)
sigmahat2
```

```
[1] 10.62
```

1.7 Compute confidence intervals by formula

```
qt(.975, n-3)

[1] 2.074

# 95% C.I. for beta2
#
beta2<-beta[3,1]
beta2

distance
 0.01438

SE.beta2<-sqrt(sigmahat2*XtXi[3,3])
SE.beta2

[1] 0.003613

qt(.975, n-3)

[1] 2.074

c(beta2-qt(.975, n-3)*SE.beta2, beta2+qt(.975, n-3)*SE.beta2)

distance distance
0.006892 0.021878
```

```
# 95% C.I. for beta1
#
beta1<-beta[2,1]
beta1

cases
1.616

SE.beta1<-sqrt(sigmahat2*XtXi[2,2])
SE.beta1

[1] 0.1707

qt(.975, n-3)

[1] 2.074

c(beta1-qt(.975, n-3)*SE.beta1, beta1+qt(.975, n-3)*SE.beta1)

cases cases
1.262 1.970
```

```
# 95% C.I. for beta0
#
beta0<-beta[1,1]
beta0

    1
2.341

SE.beta0<-sqrt(sigmahat2*XtXi[1,1])
SE.beta0

[1] 1.097

qt(.975, n-3)

[1] 2.074

c(beta0-qt(.975, n-3)*SE.beta0, beta0+qt(.975, n-3)*SE.beta0)

      1       1
0.06675 4.61571
```

1.8 Confidence intervals for model parameters

```
dfit <- lm(time ~ cases + distance, data=delivery)
confint(dfit, level=0.95)

               2.5 %  97.5 %
(Intercept) 0.066752 4.61571
cases       1.261825 1.96999
distance    0.006892 0.02188
```

1.9 Compute confidence interval for variance

```
MSE<-sum(dfit$res^2)/(n-3)
qchisq(.975,n-3)

[1] 36.78

qchisq(.025,n-3)

[1] 10.98

c((n-3)*MSE/qchisq(.975,n-3),(n-3)*MSE/qchisq(.025,n-3))

[1]  6.355 21.283
```

1.10 Compute confidence interval for the mean value of y when x=x0

```
x0<-c(1,8,275)
muhat<-t(x0) %*% beta
muhat

      [,1]
[1,] 19.22

qt(.975, n-3)

[1] 2.074

bm<-qt(.975, n-3)*sqrt(MSE*(t(x0) %*% XtXi %*%x0))
bm

     [,1]
[1,] 1.57

c(muhat-bm, muhat+bm)

[1] 17.65 20.79
```

1.11 Compute prediction interval for individual value of y when x=x0

```
x0<-c(1,8,275)
y0<-t(x0) %*% beta
y0

      [,1]
[1,] 19.22

qt(.975, n-3)

[1] 2.074

bm<-qt(.975, n-3)*sqrt(MSE*(1+t(x0) %*% XtXi %*%x0))
bm

     [,1]
[1,] 6.94

c(y0-bm, y0+bm)

[1] 12.28 26.16
```

1.12 Compute ANOVA by formula

```
sumy<-sum(delivery$time)
SST<-yty-(sumy)^2/n
SST

     [,1]
[1,] 5785

SSR<-t(beta) %*% t(X) %*% y-(sumy)^2/n
SSR

     [,1]
[1,] 5551

SSE<-SST-SSR
SSE

      [,1]
[1,] 233.7

MSR<-SSR/2
MSR

     [,1]
[1,] 2775

MSE<-SSE/(n-3)
MSE

      [,1]
[1,] 10.62

Fratio<-MSR/MSE
Fratio

      [,1]
[1,] 261.2

qf(0.95,2,n-3)

[1] 3.443

pvalue<-1-pf(Fratio,2,n-3)
pvalue

          [,1]
[1,] 4.441e-16
```

1.13 Compute ANOVA table

```
dfit <- lm(time ~ cases + distance, data=delivery)
anova(dfit)
```

```
Analysis of Variance Table

Response: time
          Df Sum Sq Mean Sq F value  Pr(>F)
cases      1   5382    5382   506.6 < 2e-16 ***
distance   1    168     168    15.8 0.00063 ***
Residuals 22    234      11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.14 Compute coefficient of determination R2 and adjusted R2

```
dfit <- lm(time ~ cases + distance, data=delivery)
summary(dfit)
```

```

Call:
lm(formula = time ~ cases + distance, data = delivery)

Residuals:
   Min     1Q Median     3Q    Max
-5.788 -0.663  0.436  1.157  7.420

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.34123    1.09673    2.13  0.04417 *
cases        1.61591    0.17073    9.46  3.3e-09 ***
distance     0.01438    0.00361    3.98  0.00063 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.26 on 22 degrees of freedom
Multiple R-squared:  0.96,Adjusted R-squared:  0.956
F-statistic:  261 on 2 and 22 DF,  p-value: 4.69e-16
```

```
R2<-SSR/SST
R2
```

```
       [,1]
[1,] 0.9596

adjR2<-1-(n-1)/(n-3)*(1-R2)
adjR2

       [,1]
[1,] 0.9559
```

1.15 t test for individual beta

```
# Hypothesis testing for beta2
#
beta2<-beta[3,1]
beta2

distance
 0.01438

SE.beta2<-sqrt(MSE*XtXi[3,3])
SE.beta2

         [,1]
[1,] 0.003613

t2<-beta2/SE.beta2
t2

      [,1]
[1,] 3.981

pvalue2<-2*(1-pt(abs(t2),n-3))
pvalue2

          [,1]
[1,] 0.0006312
```

```
# Hypothesis testing for beta1
#
beta1<-beta[2,1]
beta1

cases
1.616
```

```
SE.beta1<-sqrt(MSE*XtXi[2,2])
SE.beta1

       [,1]
[1,] 0.1707

t1<-beta1/SE.beta1
t1

      [,1]
[1,] 9.464

pvalue1<-2*(1-pt(abs(t1),n-3))
pvalue1

          [,1]
[1,] 3.255e-09
```

```
# Hypothesis testing for beta0
#
beta0<-beta[1,1]
beta0

    1
2.341

SE.beta0<-sqrt(MSE*XtXi[1,1])
SE.beta0

      [,1]
[1,] 1.097

t0<-beta0/SE.beta0
t0

      [,1]
[1,] 2.135

pvalue0<-2*(1-pt(abs(t0),n-3))
pvalue0

        [,1]
[1,] 0.04417
```

1.16 Partial F test for individual beta

```
dfit1 <- lm(time ~ cases + distance, data=delivery) # Full model
dfit2 <- lm(time ~ cases , data=delivery) # Reduced model
anova(dfit2,dfit1)

Analysis of Variance Table

Model 1: time ~ cases
Model 2: time ~ cases + distance
  Res.Df RSS Df Sum of Sq    F  Pr(>F)
1     23 402
2     22 234  1       168 15.8 0.00063 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1.17 Simultaneous confidence intervals for model parameters

```
rocket<-read.csv(file="D:/chilo/Regression 2/rocket.csv", header=T)
rocket

   observation strength   age
1            1     2159 15.50
2            2     1678 23.75
3            3     2316  8.00
4            4     2061 17.00
5            5     2208  5.50
6            6     1708 19.00
7            7     1785 24.00
8            8     2575  2.50
9            9     2358  7.50
10          10     2257 11.00
11          11     2165 13.00
12          12     2400  3.75
13          13     1780 25.00
14          14     2337  9.75
15          15     1765 22.00
16          16     2054 18.00
17          17     2414  6.00
18          18     2200 12.50
19          19     2654  2.00
20          20     1754 21.50

attach(rocket)
```

```
The following object is masked from delivery (position 3):

    observation
The following object is masked from delivery (position 4):

    observation
```

```
n<-length(age)
n
```

```
[1] 20
```

```
rocket.lm<-lm(strength~age, data=rocket)
summary(rocket.lm)
```

```

Call:
lm(formula = strength ~ age, data = rocket)

Residuals:
   Min     1Q Median     3Q    Max 
-216.0  -50.7   28.7   66.6  106.8 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2627.82      44.18    59.5  < 2e-16 ***
age           -37.15       2.89   -12.9  1.6e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.1 on 18 degrees of freedom
Multiple R-squared:  0.902,Adjusted R-squared:  0.896 
F-statistic:  165 on 1 and 18 DF,  p-value: 1.64e-10
```

```
xbar<-mean(age)
xbar
```

```
[1] 13.36
```

```
sumx<-sum(age)
sumx
```

```
[1] 267.2
```

```
sumx2<-sum(age^2)
sumx2
```

```
[1] 4678
```

```r
Sxx<-sumx2-(sumx)^2/n
Sxx

[1] 1107

beta1<-coef(rocket.lm)[2]
beta1

   age
-37.15

beta0<-coef(rocket.lm)[1]
beta0

(Intercept)
       2628

MSE<-sum((resid(rocket.lm))^2)/(n-2)
MSE

[1] 9236

SE.beta1<-sqrt(MSE/Sxx)
SE.beta1

[1] 2.889

SE.beta0<-sqrt(MSE*(1/n+xbar^2/Sxx))
SE.beta0

[1] 44.18
```

```r
#Bonferroni's method
qt(1-.025/2, n-2)

[1] 2.445

# 95% joint C.I. for beta0 and beta1
#
c(beta0-qt(1-.025/2, n-2)*SE.beta0, beta0+qt(1-.025/2, n-2)*SE.beta0)

(Intercept) (Intercept)
       2520        2736

c(beta1-qt(1-.025/2, n-2)*SE.beta1, beta1+qt(1-.025/2, n-2)*SE.beta1)

   age    age
-44.22 -30.09
```

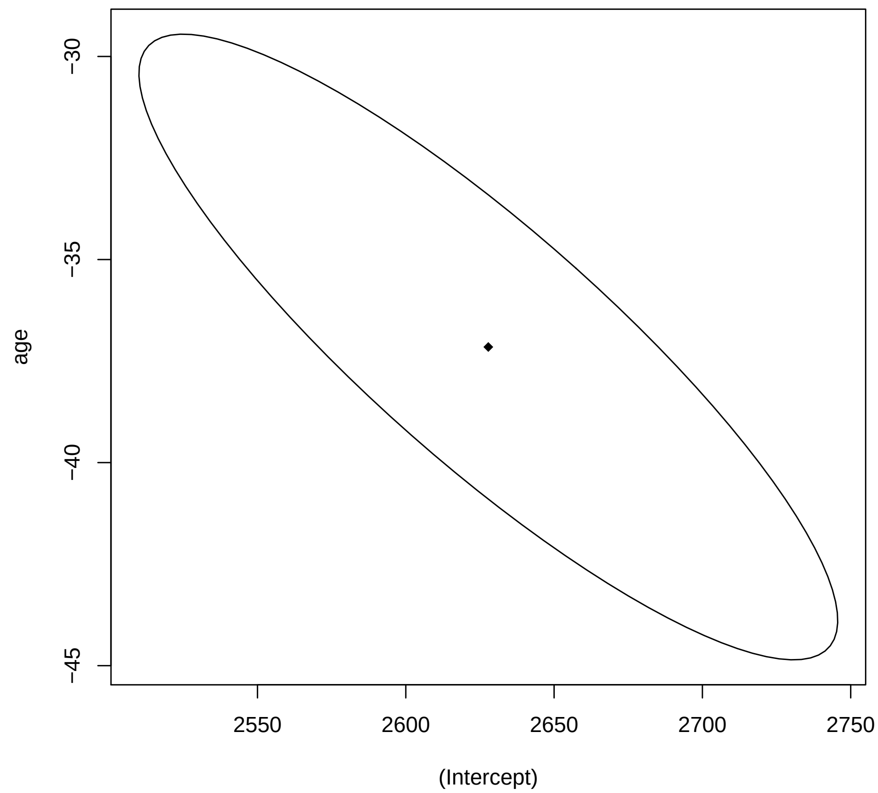
```
#Scheffe S-method
dd<-(2*qf(1-.025, 2, n-2))^0.5
dd

[1] 3.02

# 95% joint C.I. for beta0 and beta1
#
c(beta0-dd*SE.beta0, beta0+dd*SE.beta0)

(Intercept) (Intercept)
       2494        2761

c(beta1-dd*SE.beta1, beta1+dd*SE.beta1)

   age    age
-45.88 -28.43
```

```
#Confidence region for beta0 and beta1
install.packages("ellipse")

Error:  trying to use CRAN without setting a mirror

library(ellipse)

Warning:  package 'ellipse' was built under R version 3.0.3

ellipse(rocket.lm, which=c(1,2), level=.95)

       (Intercept)    age
  [1,]        2657 -35.22
  [2,]        2650 -34.75
  [3,]        2643 -34.29
  [4,]        2635 -33.84
  [5,]        2628 -33.41
  [6,]        2620 -32.99
  [7,]        2613 -32.59
  [8,]        2606 -32.20
  [9,]        2598 -31.84
 [10,]        2591 -31.49
 [11,]        2584 -31.17
 [12,]        2577 -30.88
 [13,]        2571 -30.61
 [14,]        2564 -30.36
 [15,]        2558 -30.15
 [16,]        2552 -29.96
 [17,]        2547 -29.80
```

```
[18,]         2541 -29.67
[19,]         2536 -29.57
[20,]         2532 -29.50
[21,]         2528 -29.46
[22,]         2524 -29.45
[23,]         2521 -29.47
[24,]         2518 -29.53
[25,]         2515 -29.61
[26,]         2513 -29.73
[27,]         2512 -29.87
[28,]         2511 -30.05
[29,]         2510 -30.25
[30,]         2510 -30.48
[31,]         2510 -30.74
[32,]         2511 -31.02
[33,]         2513 -31.33
[34,]         2514 -31.66
[35,]         2516 -32.02
[36,]         2519 -32.39
[37,]         2522 -32.79
[38,]         2526 -33.20
[39,]         2530 -33.62
[40,]         2534 -34.07
[41,]         2539 -34.52
[42,]         2544 -34.98
[43,]         2549 -35.46
[44,]         2555 -35.94
[45,]         2561 -36.42
[46,]         2567 -36.91
[47,]         2574 -37.40
[48,]         2581 -37.89
[49,]         2588 -38.37
[50,]         2595 -38.85
[51,]         2602 -39.32
[52,]         2609 -39.79
[53,]         2617 -40.24
[54,]         2624 -40.68
[55,]         2632 -41.11
[56,]         2639 -41.52
[57,]         2646 -41.92
[58,]         2654 -42.29
[59,]         2661 -42.64
[60,]         2668 -42.98
[61,]         2675 -43.28
[62,]         2682 -43.57
```

```
 [63,]         2688 -43.83
 [64,]         2695 -44.06
 [65,]         2701 -44.26
 [66,]         2706 -44.43
 [67,]         2712 -44.58
 [68,]         2717 -44.69
 [69,]         2722 -44.78
 [70,]         2726 -44.83
 [71,]         2730 -44.86
 [72,]         2733 -44.85
 [73,]         2736 -44.81
 [74,]         2739 -44.74
 [75,]         2741 -44.64
 [76,]         2743 -44.51
 [77,]         2744 -44.35
 [78,]         2745 -44.16
 [79,]         2746 -43.94
 [80,]         2745 -43.70
 [81,]         2745 -43.43
 [82,]         2744 -43.13
 [83,]         2742 -42.81
 [84,]         2740 -42.47
 [85,]         2738 -42.10
 [86,]         2735 -41.72
 [87,]         2732 -41.32
 [88,]         2728 -40.90
 [89,]         2724 -40.46
 [90,]         2719 -40.02
 [91,]         2714 -39.56
 [92,]         2709 -39.09
 [93,]         2704 -38.61
 [94,]         2698 -38.13
 [95,]         2692 -37.64
 [96,]         2685 -37.15
 [97,]         2678 -36.66
 [98,]         2672 -36.18
 [99,]         2665 -35.69
[100,]         2657 -35.22
```

```r
plot(ellipse(rocket.lm), type = 'l')
points(beta0,beta1,pch=18)
```


−30
−35
−40
−45
age
2550
2600
2650
2700
2750
(Intercept)