MATH3431- Practical Class Sheets 2

魏上傑

2023-03-23

```
library("faraway")
```

```
names(fat)
```

```
##  [1] "brozek"  "siri"    "density" "age"     "weight"  "height"  "adipos"
##  [8] "free"    "neck"    "chest"   "abdom"   "hip"     "thigh"   "knee"
## [15] "ankle"   "biceps"  "forearm" "wrist"
```

```
?fat
str(fat)
```

```
## 'data.frame':    252 obs. of  18 variables:
##  $ brozek : num  12.6 6.9 24.6 10.9 27.8 20.6 19 12.8 5.1 12 ...
##  $ siri   : num  12.3 6.1 25.3 10.4 28.7 20.9 19.2 12.4 4.1 11.7 ...
##  $ density: num  1.07 1.09 1.04 1.08 1.03 ...
##  $ age    : int  23 22 22 26 24 24 26 25 25 23 ...
##  $ weight : num  154 173 154 185 184 ...
##  $ height : num  67.8 72.2 66.2 72.2 71.2 ...
##  $ adipos : num  23.7 23.4 24.7 24.9 25.6 26.5 26.2 23.6 24.6 25.8 ...
##  $ free   : num  135 161 116 165 133 ...
##  $ neck   : num  36.2 38.5 34 37.4 34.4 39 36.4 37.8 38.1 42.1 ...
##  $ chest  : num  93.1 93.6 95.8 101.8 97.3 ...
##  $ abdom  : num  85.2 83 87.9 86.4 100 94.4 90.7 88.5 82.5 88.6 ...
##  $ hip    : num  94.5 98.7 99.2 101.2 101.9 ...
##  $ thigh  : num  59 58.7 59.6 60.1 63.2 66 58.4 60 62.9 63.1 ...
##  $ knee   : num  37.3 37.3 38.9 37.3 42.2 42 38.3 39.4 38.3 41.7 ...
##  $ ankle  : num  21.9 23.4 24 22.8 24 25.6 22.9 23.2 23.8 25 ...
##  $ biceps : num  32 30.5 28.8 32.4 32.2 35.7 31.9 30.5 35.9 35.6 ...
##  $ forearm: num  27.4 28.9 25.2 29.4 27.7 30.6 27.8 29 31.1 30 ...
##  $ wrist  : num  17.1 18.2 16.6 18.2 17.7 18.8 17.7 18.8 18.2 19.2 ...
```

```
head(fat)
```

```
##   brozek siri density age weight height adipos  free neck chest abdom   hip
## 1   12.6 12.3  1.0708  23 154.25  67.75   23.7 134.9 36.2  93.1  85.2  94.5
## 2    6.9  6.1  1.0853  22 173.25  72.25   23.4 161.3 38.5  93.6  83.0  98.7
## 3   24.6 25.3  1.0414  22 154.00  66.25   24.7 116.0 34.0  95.8  87.9  99.2
## 4   10.9 10.4  1.0751  26 184.75  72.25   24.9 164.7 37.4 101.8  86.4 101.2
## 5   27.8 28.7  1.0340  24 184.25  71.25   25.6 133.1 34.4  97.3 100.0 101.9
## 6   20.6 20.9  1.0502  24 210.25  74.75   26.5 167.0 39.0 104.5  94.4 107.8
##   thigh knee ankle biceps forearm wrist
## 1  59.0 37.3  21.9   32.0    27.4  17.1
## 2  58.7 37.3  23.4   30.5    28.9  18.2
## 3  59.6 38.9  24.0   28.8    25.2  16.6
## 4  60.1 37.3  22.8   32.4    29.4  18.2
## 5  63.2 42.2  24.0   32.2    27.7  17.7
## 6  66.0 42.0  25.6   35.7    30.6  18.8
```

```
summary(fat)
```

```
##      brozek            siri          density            age       
##  Min.   : 0.00   Min.   : 0.00   Min.   :0.995   Min.   :22.00  
##  1st Qu.:12.80   1st Qu.:12.47   1st Qu.:1.041   1st Qu.:35.75  
##  Median :19.00   Median :19.20   Median :1.055   Median :43.00  
##  Mean   :18.94   Mean   :19.15   Mean   :1.056   Mean   :44.88  
##  3rd Qu.:24.60   3rd Qu.:25.30   3rd Qu.:1.070   3rd Qu.:54.00  
##  Max.   :45.10   Max.   :47.50   Max.   :1.109   Max.   :81.00  
##      weight           height          adipos            free       
##  Min.   :118.5   Min.   :29.50   Min.   :18.10   Min.   :105.9  
##  1st Qu.:159.0   1st Qu.:68.25   1st Qu.:23.10   1st Qu.:131.3  
##  Median :176.5   Median :70.00   Median :25.05   Median :141.6  
##  Mean   :178.9   Mean   :70.15   Mean   :25.44   Mean   :143.7  
##  3rd Qu.:197.0   3rd Qu.:72.25   3rd Qu.:27.32   3rd Qu.:153.9  
##  Max.   :363.1   Max.   :77.75   Max.   :48.90   Max.   :240.5  
##       neck            chest            abdom             hip       
##  Min.   :31.10   Min.   : 79.30   Min.   : 69.40   Min.   : 85.0  
##  1st Qu.:36.40   1st Qu.: 94.35   1st Qu.: 84.58   1st Qu.: 95.5  
##  Median :38.00   Median : 99.65   Median : 90.95   Median : 99.3  
##  Mean   :37.99   Mean   :100.82   Mean   : 92.56   Mean   : 99.9  
##  3rd Qu.:39.42   3rd Qu.:105.38   3rd Qu.: 99.33   3rd Qu.:103.5  
##  Max.   :51.20   Max.   :136.20   Max.   :148.10   Max.   :147.7
```

```
##      thigh            knee           ankle          biceps         forearm
##  Min.   :47.20   Min.   :33.00   Min.   :19.1   Min.   :24.80   Min.   :21.00
##  1st Qu.:56.00   1st Qu.:36.98   1st Qu.:22.0   1st Qu.:30.20   1st Qu.:27.30
##  Median :59.00   Median :38.50   Median :22.8   Median :32.05   Median :28.70
##  Mean   :59.41   Mean   :38.59   Mean   :23.1   Mean   :32.27   Mean   :28.66
##  3rd Qu.:62.35   3rd Qu.:39.92   3rd Qu.:24.0   3rd Qu.:34.33   3rd Qu.:30.00
##  Max.   :87.30   Max.   :49.10   Max.   :33.9   Max.   :45.00   Max.   :34.90
##      wrist
##  Min.   :15.80
##  1st Qu.:17.60
##  Median :18.30
##  Mean   :18.23
##  3rd Qu.:18.80
##  Max.   :21.40
```

```
# check for missing values
sum(is.na(fat))
```

```
## [1] 0
```

```
# remove some variables
fat1 <- fat[,-c(2,3,8)]
```

2.1 EDA

```
length(fat1)
```

```
## [1] 15
```

```
ncol(fat1)
```

```
## [1] 15
```

```
nrow(fat1)
```

```
## [1] 252
```

```
dim(fat1)
```

```
## [1] 252  15
```

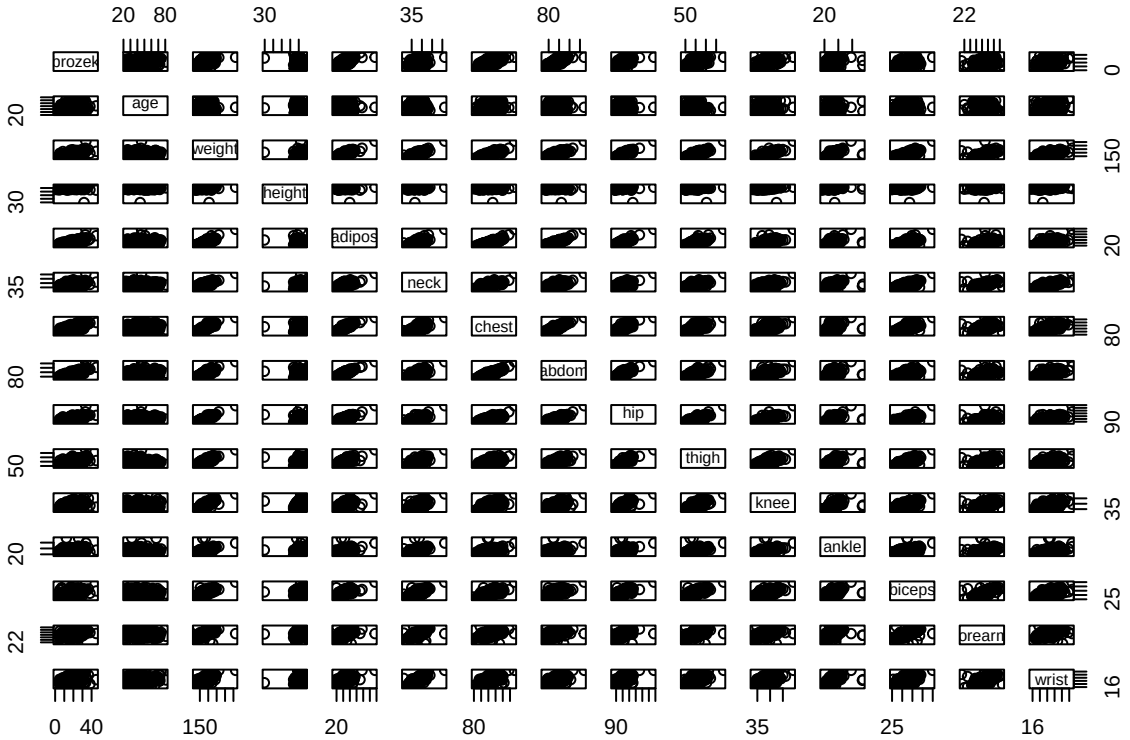
```
# 252 obs and 15 variables
```

```
cor(fat1)
```

```
##               brozek         age      weight      height      adipos      neck
## brozek   1.00000000  0.28917352  0.61315611 -0.08910641  0.72799418 0.4914889
## age      0.28917352  1.00000000 -0.01274609 -0.17164514  0.11885126 0.1135052
## weight   0.61315611 -0.01274609  1.00000000  0.30827854  0.88735216 0.8307162
## height  -0.08910641 -0.17164514  0.30827854  1.00000000 -0.02489094 0.2537099
## adipos   0.72799418  0.11885126  0.88735216 -0.02489094  1.00000000 0.7778569
## neck     0.49148893  0.11350519  0.83071622  0.25370988  0.77785691 1.0000000
## chest    0.70288516  0.17644968  0.89419052  0.13489181  0.91179865 0.7848350
## abdom    0.81370622  0.23040942  0.88799494  0.08781291  0.92388010 0.7540774
## hip      0.62569993 -0.05033212  0.94088412  0.17039426  0.88326922 0.7349579
## thigh    0.56128438 -0.20009576  0.86869354  0.14843561  0.81270609 0.6956973
## knee     0.50778587  0.01751569  0.85316739  0.28605321  0.71365983 0.6724050
## ankle    0.26678256 -0.10505810  0.61368542  0.26474369  0.50031664 0.4778924
## biceps   0.49303089 -0.04116212  0.80041593  0.20781557  0.74638418 0.7311459
## forearm  0.36327744 -0.08505555  0.63030143  0.22864922  0.55859425 0.6236603
## wrist    0.34757276  0.21353062  0.72977489  0.32206533  0.62590659 0.7448264
##               chest      abdom         hip      thigh       knee      ankle
## brozek   0.7028852 0.81370622  0.62569993  0.5612844 0.50778587  0.2667826
## age      0.1764497 0.23040942 -0.05033212 -0.2000958 0.01751569 -0.1050581
## weight   0.8941905 0.88799494  0.94088412  0.8686935 0.85316739  0.6136854
## height   0.1348918 0.08781291  0.17039426  0.1484356 0.28605321  0.2647437
## adipos   0.9117986 0.92388010  0.88326922  0.8127061 0.71365983  0.5003166
## neck     0.7848350 0.75407737  0.73495788  0.6956973 0.67240498  0.4778924
## chest    1.0000000 0.91582767  0.82941992  0.7298586 0.71949640  0.4829879
## abdom    0.9158277 1.00000000  0.87406618  0.7666239 0.73717888  0.4532227
## hip      0.8294199 0.87406618  1.00000000  0.8964098 0.82347262  0.5583868
## thigh    0.7298586 0.76662393  0.89640979  1.0000000 0.79917030  0.5397971
## knee     0.7194964 0.73717888  0.82347262  0.7991703 1.00000000  0.6116082
## ankle    0.4829879 0.45322269  0.55838682  0.5397971 0.61160820  1.0000000
## biceps   0.7279075 0.68498272  0.73927252  0.7614774 0.67870883  0.4848545
## forearm  0.5801727 0.50331609  0.54501412  0.5668422 0.55589819  0.4190500
## wrist    0.6601623 0.61983243  0.63008954  0.5586848 0.66450729  0.5661946
##               biceps     forearm     wrist
## brozek   0.49303089  0.36327744 0.3475728
## age     -0.04116212 -0.08505555 0.2135306
## weight   0.80041593  0.63030143 0.7297749
```

```
## height   0.20781557  0.22864922 0.3220653
## adipos   0.74638418  0.55859425 0.6259066
## neck     0.73114592  0.62366027 0.7448264
## chest    0.72790748  0.58017273 0.6601623
## abdom    0.68498272  0.50331609 0.6198324
## hip      0.73927252  0.54501412 0.6300895
## thigh    0.76147745  0.56684218 0.5586848
## knee     0.67870883  0.55589819 0.6645073
## ankle    0.48485454  0.41904999 0.5661946
## biceps   1.00000000  0.67825513 0.6321264
## forearm  0.67825513  1.00000000 0.5855883
## wrist    0.63212642  0.58558825 1.0000000
```

```
# pairwise scatterplots
pairs(fat1)
```



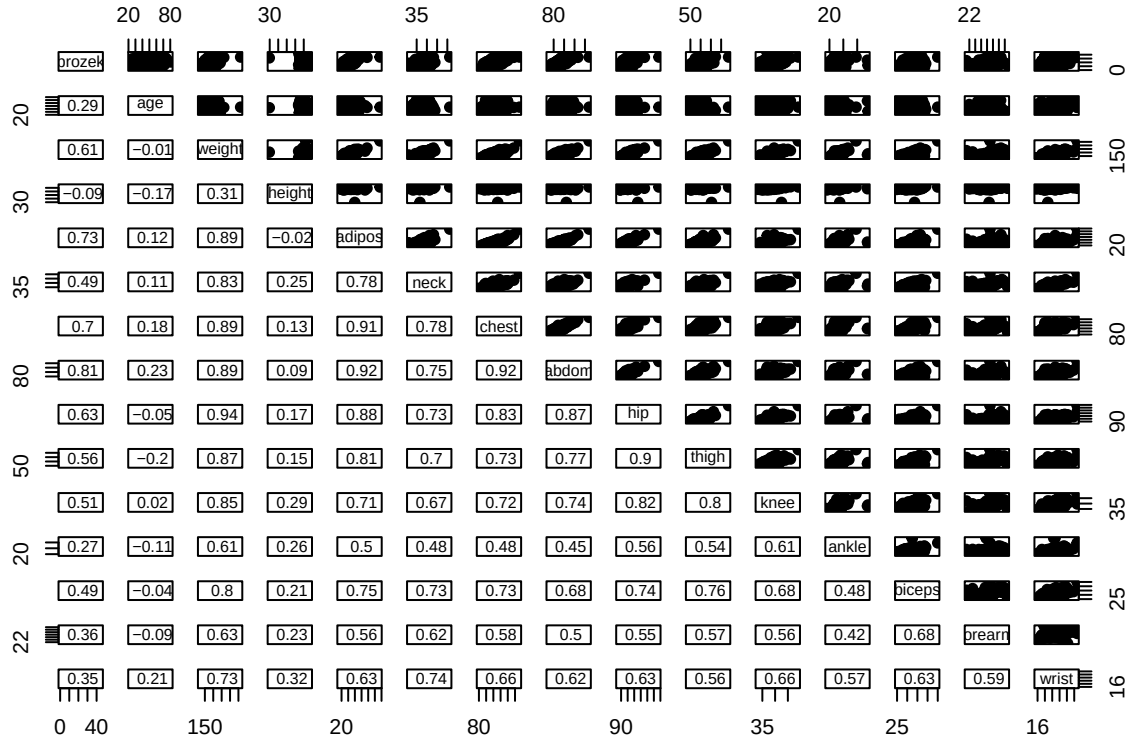

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.1     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
```

```
## v readr   2.1.4     v forcats 0.5.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
# Add correlation coefficients
# Correlation panel

# usr: A vector of the form c(x1, x2, y1, y2)
# giving the extremes of the user coordinates
# of the plotting region.

# 在函数内，可以在函数开头修改了图形参数后，用 on.exit() 函数将恢复原始图形参数作为函数退出

panel.cor <- function(x, y){
  # usr 變量用於保存當前繪圖區的坐標範圍，
  # 以便在函數結束時恢復繪圖區的原始範圍
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- round(cor(x, y), digits=2)
  txt <- paste0(" ", r)
  # 使用 text() 函數將 txt 寫入當前繪圖區的正中央位置
  # ((0.5,0.5)), 並將字元大小設置為 cex=0.8。
  text(0.5, 0.5, txt, cex = 0.8)
}


# Customize upper panel
upper.panel<-function(x, y){
  points(x,y, pch = 19)
}

# lower.panel: 用於定義下三角面板的函數
# upper.panel: 用於定義上三角面板的函數

# Create the plots
pairs(fat1,
      # 所以下三角出現相關係數
      lower.panel = panel.cor,
      # 上三角則會畫圖
```

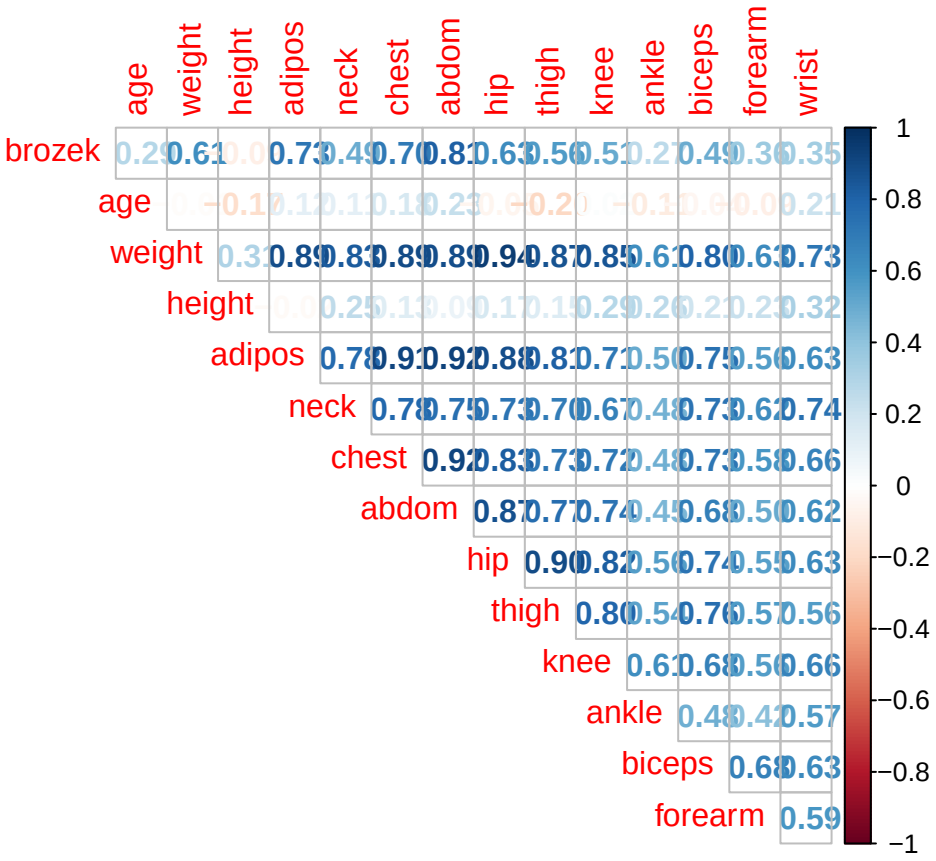
```
      upper.panel = upper.panel) %>% suppressWarnings()
```




```
library(corrplot)
```

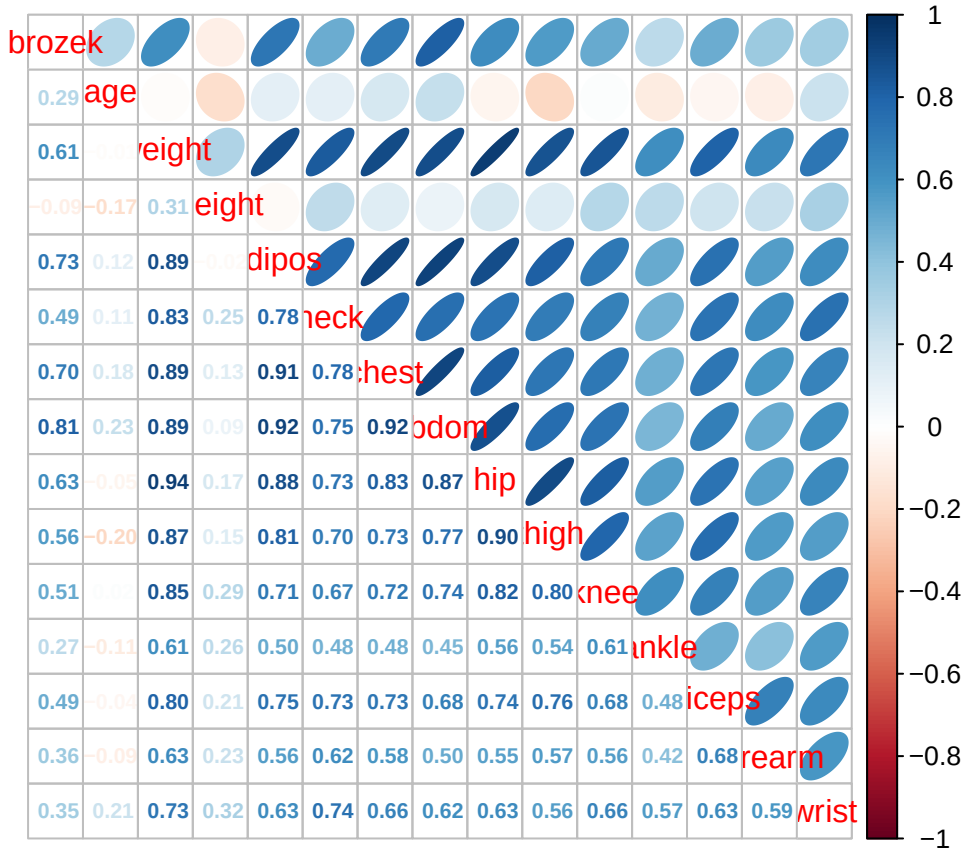
```
## corrplot 0.92 loaded
```

```
# checking correlation btw variables
corrplot(cor(fat1), method = "number", type = "upper", diag = FALSE)
```




```
# method = "number" 表示要在相关矩阵中显示相关系数值，
# type = "upper" 表示只绘制矩阵的上三角部分，
# diag = FALSE 表示不绘制对角线上的数字。
```

```
# mixed type correlation plot
corrplot.mixed(cor(fat1), upper = "ellipse", lower = "number", number.cex= .7)
```



2.2 SLR

```
reg1 <- lm(brozek~adipos, data=fat1)
summary(reg1)
```

```
##
## Call:
## lm(formula = brozek ~ adipos, data = fat1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4292  -3.4478   0.2113   3.8663  11.7826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.40508    2.36723   -8.62 7.78e-16 ***
## adipos        1.54671    0.09212   16.79  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.324 on 250 degrees of freedom
## Multiple R-squared:   0.53,  Adjusted R-squared:  0.5281
## F-statistic: 281.9 on 1 and 250 DF,  p-value: < 2.2e-16
```

```
names(reg1)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```
coef(reg1)
```

```
## (Intercept)      adipos
##  -20.405082    1.546712
```

```
reg1$coefficients
```

```
## (Intercept)      adipos
##  -20.405082    1.546712
```

2.3 Confidence and prediction intervals

```
confint(reg1, level=0.95)
```

```
##                  2.5 %    97.5 %
## (Intercept) -25.067331 -15.74283
## adipos        1.365275   1.72815
```

```
newdata <- data.frame(adipos=22)
confint(reg1, newdata=newdata, interval="confidence", level=0.95)
```

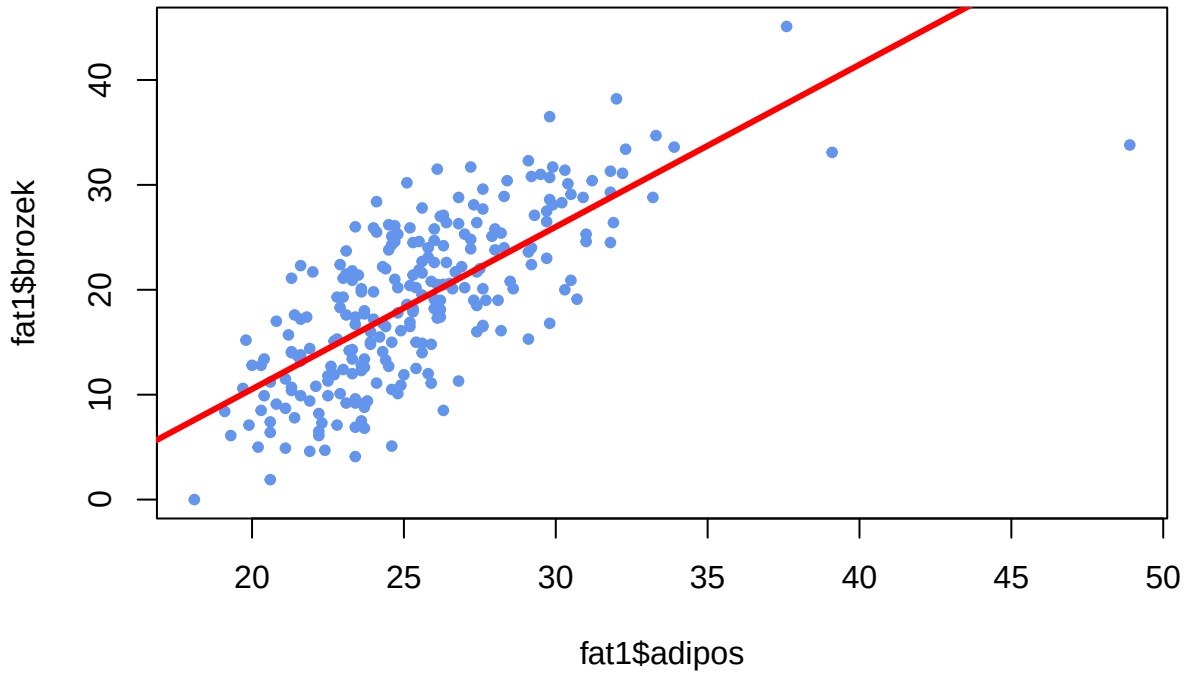
```
##                  2.5 %    97.5 %
## (Intercept) -25.067331 -15.74283
## adipos        1.365275   1.72815
```

```
predict(reg1, data.frame(adipos=22), interval = "prediction", level=0.95)
```

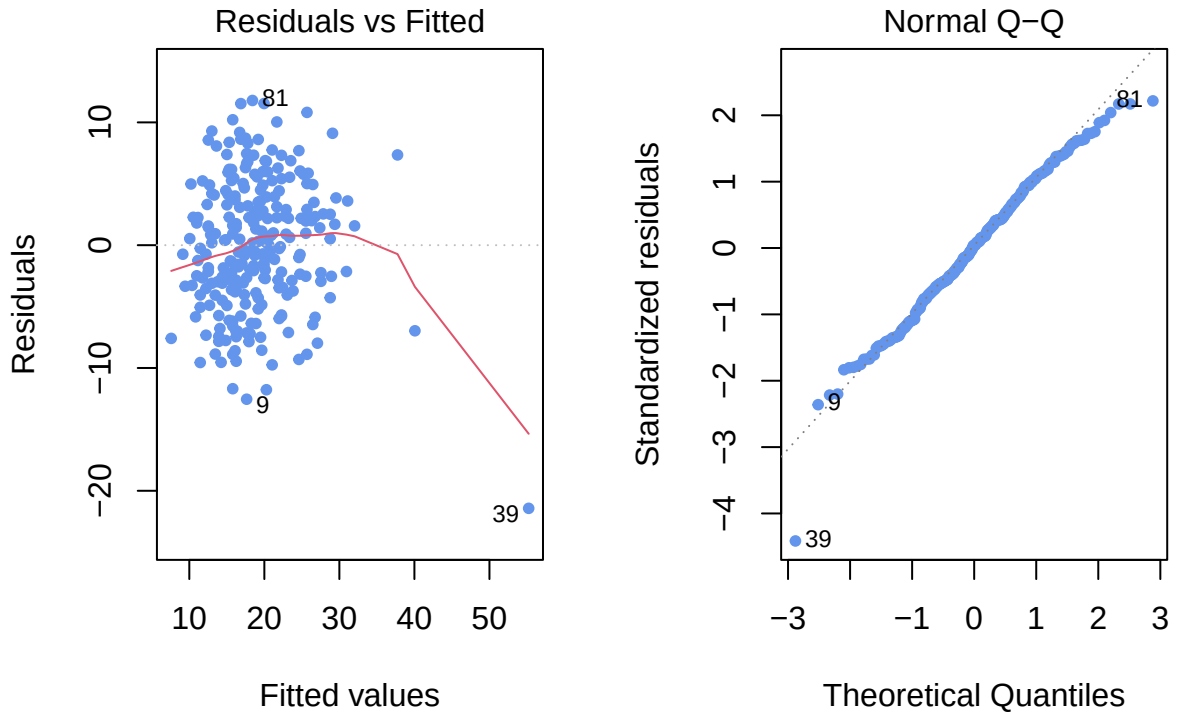
```
##        fit      lwr      upr
## 1 13.62259 3.096772 24.14841
```

2.4 Regression diagnostics

```r
plot(fat1$adipos, fat1$brozek, pch=20, col="cornflowerblue")
abline(reg1, lwd=3, col="red")
```




```r
par(mfrow=c(1,2)) # make a frame by row

# when you plot a regression model, 4 plots are generated
# we use which to choose the plot you want to present
plot(reg1, which=1, pch=20, col="cornflowerblue")
plot(reg1, which=2, pch=20, col="cornflowerblue")
```




```
par(mfrow=c(1,1)) # to return to original setting
```

2.5 MLR

```
reg2 <- lm(brozek~adipos+age, data = fat1)
reg3 <- lm(brozek~., data=fat1)
reg4 <- lm(brozek~.-age, data=fat1)
```

```
modlist <- list(reg2,reg3,reg4)
summ_modlist <- lapply(modlist, summary)
```

```
summ_modlist[[1]]$r.sq
```

```
## [1] 0.5716312
```

```
summ_modlist[[2]]$r.sq
```

```
## [1] 0.7490309
```

```
summ_modlist[[3]]$r.sq
```

```
## [1] 0.7452313
```

```
summ_modlist[[3]]$adj.r.squared
```

```
## [1] 0.7313154
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
## The following objects are masked from 'package:faraway':
##
##     logit, vif
```

```
vif(reg3)
```

```
##       age    weight    height    adipos      neck     chest     abdom       hip
##  2.250902 33.786851  2.256593 16.163444  4.430734 10.684562 13.346689 15.158277
##     thigh      knee     ankle    biceps   forearm     wrist
##  7.961508  4.828828  1.945514  3.674508  2.193390  3.379612
```

```
#Interaction terms

summary(lm(brozek~ chest*abdom, data=fat1))
```

```
##
## Call:
## lm(formula = brozek ~ chest * abdom, data = fat1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -9.9568 -3.2831  0.1953  2.8313 11.0567
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -95.206407  18.306007  -5.201 4.15e-07 ***
## chest         0.404148   0.187669   2.154 0.032243 *
## abdom         1.498560   0.204429   7.330 3.23e-12 ***
## chest:abdom  -0.006936   0.001820  -3.810 0.000175 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.332 on 248 degrees of freedom
## Multiple R-squared:  0.6913, Adjusted R-squared:  0.6876
## F-statistic: 185.1 on 3 and 248 DF,  p-value: < 2.2e-16
```

```
# main effects are automatically included
```

```
summary(lm(brozek~adipos+I(adipos^2), data=fat1))
```

```
##
## Call:
## lm(formula = brozek ~ adipos + I(adipos^2), data = fat1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8991  -3.4058  -0.1397   3.7930  11.3518
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -55.688716   7.963036  -6.993 2.46e-11 ***
## adipos        4.111738   0.561764   7.319 3.43e-12 ***
## I(adipos^2)  -0.045378   0.009814  -4.624 6.05e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.12 on 249 degrees of freedom
## Multiple R-squared:  0.5671, Adjusted R-squared:  0.5637
## F-statistic: 163.1 on 2 and 249 DF,  p-value: < 2.2e-16
```

```
summary(lm(brozek~log(adipos), data=fat1))
```

```
##
## Call:
## lm(formula = brozek ~ log(adipos), data = fat1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5263  -3.3776   0.0751   3.8273  11.4306
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -119.233      7.813  -15.26   <2e-16 ***
## log(adipos)   42.820      2.419   17.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.174 on 250 degrees of freedom
## Multiple R-squared:  0.5562, Adjusted R-squared:  0.5544
## F-statistic: 313.3 on 1 and 250 DF,  p-value: < 2.2e-16
```

2.6 Forward stepwise regression

```
library(leaps)
fwd <- regsubsets(brozek~., data=fat1, nvmax=14, method = "forward")

results <- summary(fwd)
results
```

```
## Subset selection object
## Call: regsubsets.formula(brozek ~ ., data = fat1, nvmax = 14, method = "forward")
## 14 Variables  (and intercept)
##         Forced in Forced out
## age         FALSE      FALSE
## weight      FALSE      FALSE
## height      FALSE      FALSE
## adipos      FALSE      FALSE
## neck        FALSE      FALSE
## chest       FALSE      FALSE
## abdom       FALSE      FALSE
```

```
## hip         FALSE      FALSE
## thigh       FALSE      FALSE
## knee        FALSE      FALSE
## ankle       FALSE      FALSE
## biceps      FALSE      FALSE
## forearm     FALSE      FALSE
## wrist       FALSE      FALSE
## 1 subsets of each size up to 14
## Selection Algorithm: forward
##           age weight height adipos neck chest abdom hip thigh knee ankle biceps
## 1  ( 1 )  " " " "    " "    " "    " "  " "   "*"   " " " " " "  " "   " "   " "
## 2  ( 1 )  " " "*"    " "    " "    " "  " "   "*"   " " " " " "  " "   " "   " "
## 3  ( 1 )  " " "*"    " "    " "    " "  " "   "*"   " " " " " "  " "   " "   " "
## 4  ( 1 )  " " "*"    " "    " "    " "  " "   "*"   " " " " " "  " "   " "   " "
## 5  ( 1 )  " " "*"    " "    " "    "*"  " "   "*"   " " " " " "  " "   " "   " "
## 6  ( 1 )  "*" "*"    " "    " "    "*"  " "   "*"   " " " " " "  " "   " "   " "
## 7  ( 1 )  "*" "*"    " "    " "    "*"  " "   "*"   " " "*" " "  " "   " "   " "
## 8  ( 1 )  "*" "*"    " "    " "    "*"  " "   "*"   "*" "*" " "  " "   " "   " "
## 9  ( 1 )  "*" "*"    " "    " "    "*"  " "   "*"   "*" "*" " "  " "   " "   "*"
## 10  ( 1 ) "*" "*"    " "    " "    "*"  " "   "*"   "*" "*" " "  " "   "*"   "*"
## 11  ( 1 ) "*" "*"    "*"    " "    "*"  " "   "*"   "*" "*" " "  " "   "*"   "*"
## 12  ( 1 ) "*" "*"    "*"    " "    "*"  "*"   "*"   "*" "*" " "  " "   "*"   "*"
## 13  ( 1 ) "*" "*"    "*"    "*"    "*"  "*"   "*"   "*" "*" " "  " "   "*"   "*"
## 14  ( 1 ) "*" "*"    "*"    "*"    "*"  "*"   "*"   "*" "*" "*"  "*"   "*"   "*"
##           forearm wrist
## 1  ( 1 )  " "     " "
## 2  ( 1 )  " "     " "
## 3  ( 1 )  " "     "*"
## 4  ( 1 )  "*"     "*"
## 5  ( 1 )  "*"     "*"
## 6  ( 1 )  "*"     "*"
## 7  ( 1 )  "*"     "*"
## 8  ( 1 )  "*"     "*"
## 9  ( 1 )  "*"     "*"
## 10  ( 1 ) "*"     "*"
## 11  ( 1 ) "*"     "*"
## 12  ( 1 ) "*"     "*"
## 13  ( 1 ) "*"     "*"
## 14  ( 1 ) "*"     "*"
```

```r
RSS <- results$rss
r2 <- results$rsq
Cp <- results$cp
BIC <- results$bic
Adj_r2 <- results$adjr2

# combine them into a matrix
criteria_values <- cbind(RSS, r2, Cp, BIC, Adj_r2)
```

```r
which.min(Cp)
```

```
## [1] 8
```

```r
which.min(BIC)
```

```
## [1] 4
```

```r
which.max(Adj_r2)
```

```
## [1] 8
```

```r
coef(fwd, 4)
```

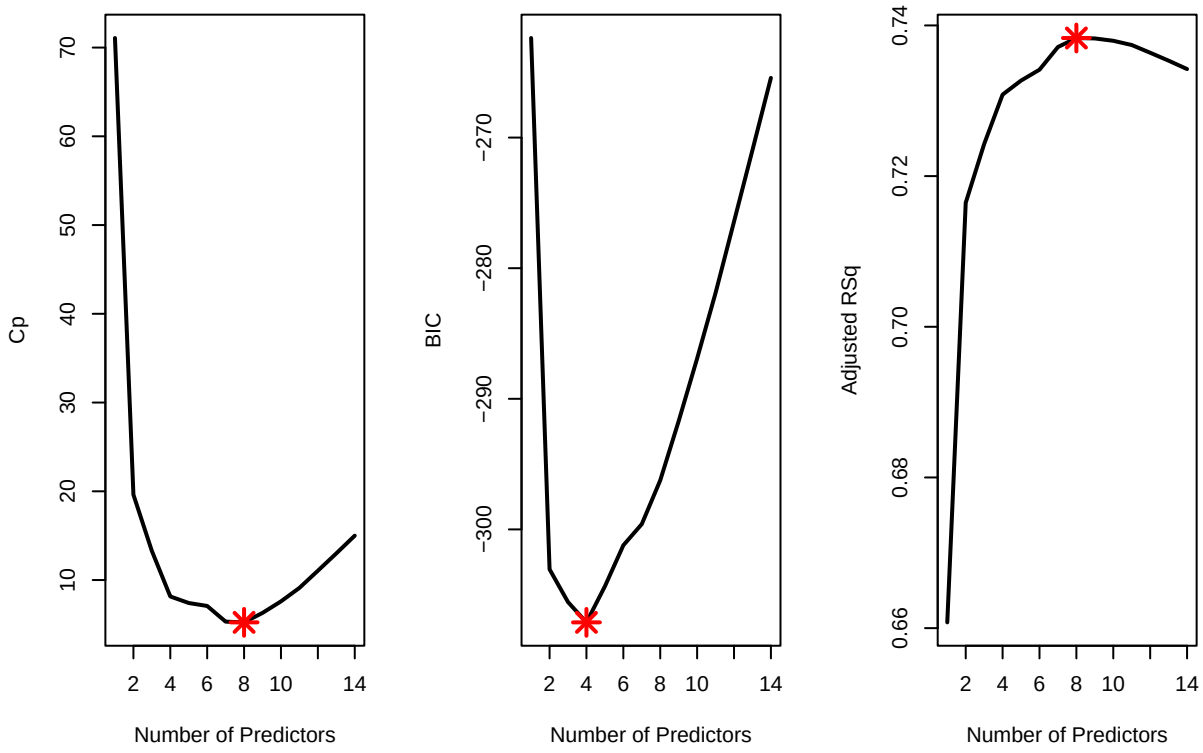
```
## (Intercept)      weight       abdom     forearm       wrist
## -31.2967858  -0.1255654   0.9213725   0.4463824  -1.3917662
```

```r
# Looking at the summary above, the model with 4 predictors (minimum BIC)
# includes the predictors weight, abdom, forearm and wrist.
coef(fwd, 8)
```

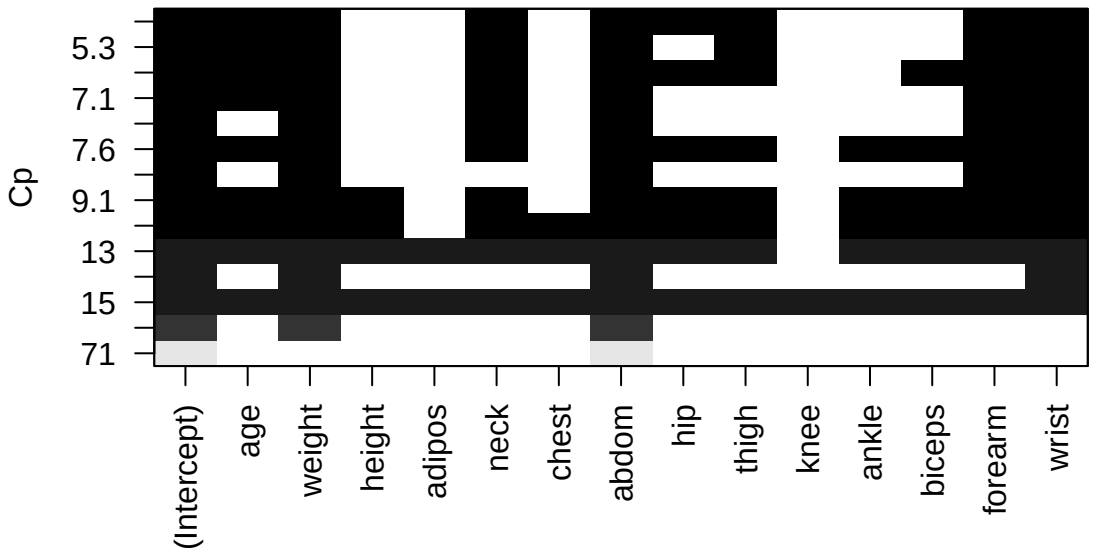
```
##  (Intercept)          age       weight         neck        abdom          hip
## -20.06213373   0.05921577  -0.08413521  -0.43189267   0.87720667  -0.18641032
##        thigh      forearm        wrist
##   0.28644340   0.48254563  -1.40486912
```

```r
# The model with 8 predictors (minimum Cp and maximum adjusted R-squared)
# includes additionally the predictors age, neck, hip and thigh.

par(mfrow = c(1, 3))
```

```
plot(x=Cp, xlab = "Number of Predictors", ylab = "Cp", type = 'l', lwd = 2)
points(x=8, y=Cp[8], col = "red", cex = 2, pch = 8, lwd = 2)

plot(x=BIC, xlab = "Number of Predictors", ylab = "BIC", type = 'l', lwd = 2)
points(x=4, y=BIC[4], col = "red", cex = 2, pch = 8, lwd = 2)

plot(x=Adj_r2, xlab = "Number of Predictors", ylab = "Adjusted RSq",
     type = "l", lwd = 2)
points(x=8, y=Adj_r2[8],  col = "red", cex = 2, pch = 8, lwd = 2)
```




```
# visualize the models obtained by Cp
par(mfrow=c(1,1))
plot(fwd, scale="Cp")
```



2.7 Best subset and Backward selection

```
best <- regsubsets(brozek~., data=fat1, nvmax=14)
bwd <- regsubsets(brozek~., data=fat1, nvmax=14, method = "backward")
```

```
which.min(summary(best)$cp)
```

```
## [1] 8
```

```
which.min(summary(best)$bic)
```

```
## [1] 4
```

```
which.max(summary(best)$adjr2)
```

```
## [1] 8
```

```
which.min(summary(bwd)$cp)
```

```
## [1] 8
```

```
which.min(summary(bwd)$bic)
```

```
## [1] 4
```

```
which.max(summary(bwd)$adjr2)
```

```
## [1] 8
```

```
# Yes, the three optimal models (under each of the criteria Cp, BIC and
# adj-R-squared) for each of forward stepwise, backward stepwise and
# best subset selections all have the same number of predictors.
```

```
# Cp
coef(fwd,8)
```

```
##  (Intercept)          age       weight         neck        abdom          hip
## -20.06213373   0.05921577  -0.08413521  -0.43189267   0.87720667  -0.18641032
##        thigh      forearm        wrist
##   0.28644340   0.48254563  -1.40486912
```

```
coef(best,8)
```

```
##  (Intercept)          age       weight         neck        abdom          hip
## -20.06213373   0.05921577  -0.08413521  -0.43189267   0.87720667  -0.18641032
##        thigh      forearm        wrist
##   0.28644340   0.48254563  -1.40486912
```

```
coef(bwd, 8)
```

```
##  (Intercept)          age       weight         neck        abdom          hip
## -20.06213373   0.05921577  -0.08413521  -0.43189267   0.87720667  -0.18641032
##        thigh      forearm        wrist
##   0.28644340   0.48254563  -1.40486912
```

```
# BIC
coef(fwd, 4)
```

```
## (Intercept)      weight       abdom     forearm       wrist
## -31.2967858  -0.1255654   0.9213725   0.4463824  -1.3917662
```

```r
coef(best, 4)
```

```
## (Intercept)      weight       abdom     forearm       wrist
## -31.2967858  -0.1255654   0.9213725   0.4463824  -1.3917662
```

```r
coef(bwd, 4)
```

```
## (Intercept)      weight       abdom     forearm       wrist
## -31.2967858  -0.1255654   0.9213725   0.4463824  -1.3917662
```

```r
# In addition,  the predictors are also the same.

# adjusted R squared is not needed because it has also 8 predictors
```