hw6

Xinrui Hu

2023-03-18

```
library (faraway)
```

Question 1: Fit a regression model with the brozek variable (percent of body fat) as a response and the following 13 predictors: age, weight, height, neck, chest, abdom, hip, thigh, knee, ankle, biceps, forearm, and wrist.

```
model <- lm(brozek ~ age + weight + height + neck + chest + abdom + hip + thigh + knee + ankle + biceps
summary(model)
```

```
##
## Call:
## lm(formula = brozek ~ age + weight + height + neck + chest +
##     abdom + hip + thigh + knee + ankle + biceps + forearm + wrist,
##     data = fat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.264  -2.572  -0.097   2.898   9.327
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.29255   16.06992  -0.952  0.34225
## age           0.05679    0.02996   1.895  0.05929 .
## weight       -0.08031    0.04958  -1.620  0.10660
## height       -0.06460    0.08893  -0.726  0.46830
## neck         -0.43754    0.21533  -2.032  0.04327 *
## chest        -0.02360    0.09184  -0.257  0.79740
## abdom         0.88543    0.08008  11.057  < 2e-16 ***
## hip          -0.19842    0.13516  -1.468  0.14341
## thigh         0.23190    0.13372   1.734  0.08418 .
## knee         -0.01168    0.22414  -0.052  0.95850
## ankle         0.16354    0.20514   0.797  0.42614
## biceps        0.15280    0.15851   0.964  0.33605
## forearm       0.43049    0.18445   2.334  0.02044 *
## wrist        -1.47654    0.49552  -2.980  0.00318 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.988 on 238 degrees of freedom
## Multiple R-squared:  0.749,  Adjusted R-squared:  0.7353
## F-statistic: 54.63 on 13 and 238 DF,  p-value: < 2.2e-16
```

Question 2 Calculate pairwse correlations between the predictors (correlation matrix). Make sure to remove the outcome variable from the matrix.

```
var1Removed <- fat[,-1]
corMatrix <- cor(var1Removed)
corMatrix <- round(corMatrix,2)
corMatrix
```

```
##          siri density   age weight height adipos  free  neck chest abdom   hip
## siri     1.00   -0.99  0.29   0.61  -0.09   0.73  0.02  0.49  0.70  0.81  0.63
## density -0.99    1.00 -0.28  -0.59   0.10  -0.71 -0.01 -0.47 -0.68 -0.80 -0.61
## age      0.29   -0.28  1.00  -0.01  -0.17   0.12 -0.24  0.11  0.18  0.23 -0.05
## weight   0.61   -0.59 -0.01   1.00   0.31   0.89  0.79  0.83  0.89  0.89  0.94
## height  -0.09    0.10 -0.17   0.31   1.00  -0.02  0.49  0.25  0.13  0.09  0.17
## adipos   0.73   -0.71  0.12   0.89  -0.02   1.00  0.55  0.78  0.91  0.92  0.88
## free     0.02   -0.01 -0.24   0.79   0.49   0.55  1.00  0.68  0.59  0.50  0.70
## neck     0.49   -0.47  0.11   0.83   0.25   0.78  0.68  1.00  0.78  0.75  0.73
## chest    0.70   -0.68  0.18   0.89   0.13   0.91  0.59  0.78  1.00  0.92  0.83
## abdom    0.81   -0.80  0.23   0.89   0.09   0.92  0.50  0.75  0.92  1.00  0.87
## hip      0.63   -0.61 -0.05   0.94   0.17   0.88  0.70  0.73  0.83  0.87  1.00
## thigh    0.56   -0.55 -0.20   0.87   0.15   0.81  0.68  0.70  0.73  0.77  0.90
## knee     0.51   -0.50  0.02   0.85   0.29   0.71  0.70  0.67  0.72  0.74  0.82
## ankle    0.27   -0.26 -0.11   0.61   0.26   0.50  0.58  0.48  0.48  0.45  0.56
## biceps   0.49   -0.49 -0.04   0.80   0.21   0.75  0.65  0.73  0.73  0.68  0.74
## forearm  0.36   -0.35 -0.09   0.63   0.23   0.56  0.55  0.62  0.58  0.50  0.55
## wrist    0.35   -0.33  0.21   0.73   0.32   0.63  0.67  0.74  0.66  0.62  0.63
##          thigh  knee ankle biceps forearm wrist
## siri      0.56  0.51  0.27   0.49    0.36  0.35
## density  -0.55 -0.50 -0.26  -0.49   -0.35 -0.33
## age      -0.20  0.02 -0.11  -0.04   -0.09  0.21
## weight    0.87  0.85  0.61   0.80    0.63  0.73
## height    0.15  0.29  0.26   0.21    0.23  0.32
## adipos    0.81  0.71  0.50   0.75    0.56  0.63
## free      0.68  0.70  0.58   0.65    0.55  0.67
## neck      0.70  0.67  0.48   0.73    0.62  0.74
## chest     0.73  0.72  0.48   0.73    0.58  0.66
## abdom     0.77  0.74  0.45   0.68    0.50  0.62
## hip       0.90  0.82  0.56   0.74    0.55  0.63
## thigh     1.00  0.80  0.54   0.76    0.57  0.56
## knee      0.80  1.00  0.61   0.68    0.56  0.66
## ankle     0.54  0.61  1.00   0.48    0.42  0.57
## biceps    0.76  0.68  0.48   1.00    0.68  0.63
## forearm   0.57  0.56  0.42   0.68    1.00  0.59
## wrist     0.56  0.66  0.57   0.63    0.59  1.00
```

Question 3 Evaluate the collinearity by calculating the condition numbers of the eigenvalues. Are there any condition numbers above 30?

```
x <- model.matrix(model)[,-1]
e <- eigen(t(x) %*% x)
e$val
```

```
##  [1] 1.959256e+07 6.418499e+04 3.059739e+04 5.704341e+03 2.803947e+03
##  [6] 1.934715e+03 1.030340e+03 6.376692e+02 5.280964e+02 4.318186e+02
## [11] 3.763758e+02 2.723663e+02 6.345357e+01
```

```
sqrt(e$val[1] / e$val)
```

```
##  [1]   1.00000  17.47144  25.30482  58.60610  83.59121 100.63222 137.89717
```

```
##  [8] 175.28623 192.61449 213.00748 228.15747 268.20620 555.67072
```

```
# There are condition value above 30
```

Question 4 Calculate the variance inflation factor for each predictor. Which inflation factors are greater than 10?

```
rsq <- summary(lm(x[,1] ~ x[,-1]))$r.squared
rsq
```

```
## [1] 0.5556445
```

```
1 / ( 1 - rsq)
```

```
## [1] 2.25045
```

```
require(faraway)
vif(x)
```

```
##       age    weight    height      neck     chest     abdom       hip     thigh
##  2.250450 33.509320  1.674591  4.324463  9.460877 11.767073 14.796520  7.777865
##      knee     ankle    biceps   forearm     wrist
##  4.612147  1.907961  3.619744  2.192492  3.377515
```

```
# weight and abdom inflation factor are greater than 10
```

Question 5 Fit a model with brozek as the response and age, weight, and height as predictors. Show the summary:

```
model1 <- lm(brozek ~ age + weight + height, fat)
summary(model1)
```

```
##
## Call:
## lm(formula = brozek ~ age + weight + height, data = fat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.0023  -4.1099  -0.0371   3.4873  14.4576
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.72142    6.92955   2.557   0.0111 *
## age          0.15583    0.02739   5.690 3.57e-08 ***
## weight       0.18373    0.01216  15.107  < 2e-16 ***
## height      -0.55099    0.09904  -5.563 6.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.382 on 248 degrees of freedom
## Multiple R-squared:  0.5236, Adjusted R-squared:  0.5179
## F-statistic: 90.87 on 3 and 248 DF,  p-value: < 2.2e-16
```

Question 6 Evaluate the collinearity by calculating the condition numbers of the eigenvalues. Are there any condition numbers above 30?

```
x1<- model.matrix(model1)[,-1]
e1 <- eigen(t(x1) %*% x1)
e1$val
```

```
## [1] 10001051.85    54778.52    19455.70
```

```
sqrt(e1$val[1] / e1$val)
```

```
## [1]  1.00000 13.51194 22.67250
```

```
# no condition number larger than 30
```

Question 7 Calculate the variance inflation factor for each predictor. Which inflation factors are greater than 10?

```
rsq1 <- summary(lm(x[,1] ~ x[,-1]))$r.squared
rsq1
```

```
## [1] 0.5556445
```

```
1 / ( 1 - rsq1)
```

```
## [1] 2.25045
```

```
require(faraway)
vif(x1)
```

```
##      age   weight   height
## 1.032253 1.107050 1.140470
```

```
# no factor greater than 10
```