W04-02-REG-df2015na Homework

문제 정답

모형	키예측식
Mlm	y = 104.86782 + 5.31866gndM - 0.06027age + 0.39190wt -0.38732wa + 2.12427hdwd + 1.57590ftln - 1.08832ftwd + 0.55505bldAB - 0.31038bldB -0.13907bldO + 0.18510lftY - 0.32792smkY - 0.43018alcY
Mstep	y = 105.41125 + 5.27676gndM -0.06226age + 0.39174wt - 0.38821wa + 2.19044hdln + 1.49213ftln - 1.03522ftwd

모형	RSQ	MSE	MAE	AIC
Mlm	30.8802	25.30651	3.66709	645.15
Mstep	0.879	21.66815	3.601111	633.84

모형	gnd	age	ht	wt	wa	hdln	hdwd	ftln	ftwd	bld	lft	smk	alc	yhlm	reslm	yhstep	resstep
본인	F	23	154	51.4	68	16	9	20	9	AB	0	0	0	153.7971	0.202867	153.2895	0.7105283
A	M	53	169	67	86	18	10.5	25	10	B	0	0	1	166.2341	2.765890	166.6282	2.3718285
B	M	20	174	75	80	19	11	26	11	B	0	0	1	176.3076	-2.307604	176.7933	-2.7933482
C	F	52	155	60	72	17.5	9	25	9.5	AB	0	0	0	164.3920	-9.392002	163.5287	-8.5287394

패키지

```
suppressWarnings(suppressMessages(library(tidyverse)))
suppressWarnings(suppressMessages(library(skimr)))
suppressWarnings(suppressMessages(library(naniar)))
suppressWarnings(suppressMessages(library(gridExtra)))
suppressWarnings(suppressMessages(library(scales)))
suppressWarnings(suppressMessages(library(dplyr)))
suppressWarnings(suppressMessages(library(caret)))
suppressWarnings(suppressMessages(library(MASS)))
```

읽기

```
DF <- as.data.frame(read_csv('D:/Github/Statics/DataMining/0404/df2015na.csv'))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   gnd = col_character(),
##   age = col_double(),
##   ht = col_double(),
##   wt = col_double(),
##   wa = col_double(),
##   hdln = col_double(),
##   hdwd = col_double(),
##   ftln = col_double(),
##   ftwd = col_double(),
##   bld = col_character(),
##   lft = col_double(),
##   smk = col_double(),
##   alc = col_double()
## )
```

```
SC <- as.data.frame(read_csv('D:/Github/Statics/DataMining/0404/df2015na-sc.csv'))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   gnd = col_character(),
##   age = col_double(),
##   ht = col_double(),
##   wt = col_double(),
##   wa = col_double(),
##   hdln = col_double(),
##   hdwd = col_double(),
##   ftln = col_double(),
##   ftwd = col_double(),
##   bld = col_character(),
##   lft = col_double(),
##   smk = col_double(),
##   alc = col_double()
## )
```

```
head(DF)
```

```
##   gnd age    ht   wt   wa hdln hdwd ftln ftwd bld lft smk alc
## 1   M  42 165.0 79.0 96.4 17.9  8.6 24.5  9.6   0   0   0   1
## 2   M  23 187.5 77.8 76.6 20.4  8.3 28.9 10.6   0   0   1   0
## 3   F  32 161.7 59.6 83.0 17.5  7.7 23.3 10.4   A   0   0   0
## 4   F  30 162.0 52.0 65.8 16.4  6.6 23.7  8.4   B   0   0   0
## 5   F  NA 160.0 58.9 75.0 17.4  7.7 24.1  9.4   A   0   0   0
## 6   M  26 179.0 76.3 83.8 19.0  8.6 25.7 10.7   B   0   0   0
```

변수 조정

```
sapply(DF, class)
```

```
##         gnd         age          ht          wt          wa        hdln
## "character"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
##        hdwd        ftln        ftwd         bld         lft         smk
##   "numeric"   "numeric"   "numeric" "character"   "numeric"   "numeric"
##         alc
##   "numeric"
```

```
sapply(SC, class)
```

```
##         gnd         age          ht          wt          wa        hdln
## "character"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
##        hdwd        ftln        ftwd         bld         lft         smk
##   "numeric"   "numeric"   "numeric" "character"   "numeric"   "numeric"
##         alc
##   "numeric"
```

```
DF <-mutate(DF,
            gnd = factor(gnd),
            bld = factor(bld),
            lft = factor(lft, labels = c('N', 'Y')),
            smk = factor(smk, labels = c('N', 'Y')),
            alc = factor(alc, labels = c('N', 'Y')))

str(DF)
```

```
## 'data.frame':    300 obs. of  13 variables:
##  $ gnd : Factor w/ 2 levels "F","M": 2 2 1 1 1 2 2 2 2 1 ...
##  $ age : num  42 23 32 30 NA 26 20 24 18 58 ...
##  $ ht  : num  165 188 162 162 160 ...
##  $ wt  : num  79 77.8 59.6 52 58.9 76.3 61.4 63 65.2 58.3 ...
##  $ wa  : num  96.4 76.6 83 65.8 75 83.8 70.1 72.4 73.5 83.1 ...
##  $ hdln: num  17.9 20.4 17.5 16.4 17.4 19 19.8 18.5 20.1 15.6 ...
##  $ hdwd: num  8.6 8.3 7.7 6.6 7.7 8.6 7.8 7.9 8 7.6 ...
##  $ ftln: num  24.5 28.9 23.3 23.7 24.1 25.7 26.5 25.7 25.8 22.1 ...
##  $ ftwd: num  9.6 10.6 10.4 8.4 9.4 10.7 9.6 10.2 10.3 8.7 ...
##  $ bld : Factor w/ 4 levels "A","AB","B","O": 4 4 1 3 1 3 2 1 4 3 ...
##  $ lft : Factor w/ 2 levels "N","Y": 1 1 1 1 1 1 1 1 1 1 ...
##  $ smk : Factor w/ 2 levels "N","Y": 1 2 1 1 1 1 1 1 1 1 ...
##  $ alc : Factor w/ 2 levels "N","Y": 2 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   gnd = col_character(),
##   ..   age = col_double(),
##   ..   ht = col_double(),
##   ..   wt = col_double(),
##   ..   wa = col_double(),
##   ..   hdln = col_double(),
##   ..   hdwd = col_double(),
##   ..   ftln = col_double(),
##   ..   ftwd = col_double(),
##   ..   bld = col_character(),
##   ..   lft = col_double(),
##   ..   smk = col_double(),
##   ..   alc = col_double()
##   .. )
```

```
head(SC)
```

```
##   gnd age  ht   wt wa hdln hdwd ftln ftwd bld lft smk alc
## 1   F  23 154 51.4 68 16.0  9.0   20  9.0  AB   0   0   0
## 2   M  53 169 67.0 86 18.0 10.5   25 10.0   B   0   0   1
## 3   M  20 174 75.0 80 19.0 11.0   26 11.0   B   0   0   1
## 4   F  52 155 60.0 72 17.5  9.0   25  9.5  AB   0   0   0
```

```
SC <-mutate(SC,
            gnd = factor(gnd),
            bld = factor(bld),
            lft = factor(lft, labels = c('N')),
            smk = factor(smk, labels = c('N')),
            alc = factor(alc, labels = c('N', 'Y')))
levels(SC$lft)<-append(levels(SC$lft), 'Y')
levels(SC$smk)<-append(levels(SC$smk), 'Y')
levels(SC$bld)<-append(levels(SC$bld), c('A','O'))
str(SC)
```

```
## 'data.frame':    4 obs. of  13 variables:
##  $ gnd : Factor w/ 2 levels "F","M": 1 2 2 1
##  $ age : num  23 53 20 52
##  $ ht  : num  154 169 174 155
##  $ wt  : num  51.4 67 75 60
##  $ wa  : num  68 86 80 72
##  $ hdln: num  16 18 19 17.5
##  $ hdwd: num  9 10.5 11 9
##  $ ftln: num  20 25 26 25
##  $ ftwd: num  9 10 11 9.5
##  $ bld : Factor w/ 4 levels "AB","B","A","O": 1 2 2 1
##  $ lft : Factor w/ 2 levels "N","Y": 1 1 1 1
##  $ smk : Factor w/ 2 levels "N","Y": 1 1 1 1
##  $ alc : Factor w/ 2 levels "N","Y": 1 2 2 1
##  - attr(*, "spec")=
##   .. cols(
##   ..   gnd = col_character(),
##   ..   age = col_double(),
##   ..   ht = col_double(),
##   ..   wt = col_double(),
##   ..   wa = col_double(),
##   ..   hdln = col_double(),
##   ..   hdwd = col_double(),
##   ..   ftln = col_double(),
##   ..   ftwd = col_double(),
##   ..   bld = col_character(),
##   ..   lft = col_double(),
##   ..   smk = col_double(),
##   ..   alc = col_double()
##   .. )
```

결측

```
skimr::skim(DF)
```

Data summary

Name	DF
Number of rows	300
Number of columns	13

Column type frequency:	
factor	5
numeric	8
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
gnd	0	1.00	FALSE	2	M: 155, F: 145
bld	0	1.00	FALSE	4	B: 110, A: 84, O: 68, AB: 38
lft	5	0.98	FALSE	2	N: 283, Y: 12
smk	2	0.99	FALSE	2	N: 237, Y: 61
alc	6	0.98	FALSE	2	N: 256, Y: 38

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
age	7	0.98	29.91	13.33	15.0	19.0	26.00	37.00	68.0	
ht	0	1.00	165.64	9.03	144.3	158.0	165.85	172.83	190.5	
wt	3	0.99	63.77	12.38	39.9	55.6	61.80	71.20	106.1	
wa	5	0.98	78.23	9.72	59.0	71.0	77.20	84.15	111.6	
hdln	0	1.00	17.57	1.09	14.9	16.8	17.50	18.30	20.8	
hdwd	1	1.00	7.80	0.55	6.4	7.4	7.80	8.20	9.2	
ftln	4	0.99	24.37	1.51	19.9	23.3	24.30	25.50	28.9	
ftwd	1	1.00	9.68	0.70	8.0	9.2	9.70	10.15	12.2	

```
naniar::miss_var_summary(DF)
```

```
## # A tibble: 13 x 3
##    variable n_miss pct_miss
##    <chr>     <int>    <dbl>
##  1 age           7    2.33
##  2 alc           6    2
##  3 wa            5    1.67
##  4 lft           5    1.67
##  5 ftln          4    1.33
##  6 wt            3    1
##  7 smk           2    0.667
##  8 hdwd          1    0.333
##  9 ftwd          1    0.333
## 10 gnd           0    0
## 11 ht            0    0
## 12 hdln          0    0
## 13 bld           0    0
```

```
skimr::skim(SC)
```

Data summary

Name	SC
Number of rows	4
Number of columns	13
_____	
Column type frequency:	
factor	5
numeric	8
_____	
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
gnd	0	1	FALSE	2	F: 2, M: 2
bld	0	1	FALSE	2	AB: 2, B: 2, A: 0, O: 0
lft	0	1	FALSE	1	N: 4, Y: 0
smk	0	1	FALSE	1	N: 4, Y: 0
alc	0	1	FALSE	2	N: 2, Y: 2

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
age	0	1	37.00	17.94	20.0	22.25	37.50	52.25	53	
ht	0	1	163.00	10.03	154.0	154.75	162.00	170.25	174	
wt	0	1	63.35	10.05	51.4	57.85	63.50	69.00	75	
wa	0	1	76.50	8.06	68.0	71.00	76.00	81.50	86	
hdln	0	1	17.62	1.25	16.0	17.12	17.75	18.25	19	
hdwd	0	1	9.88	1.03	9.0	9.00	9.75	10.62	11	
ftln	0	1	24.00	2.71	20.0	23.75	25.00	25.25	26	
ftwd	0	1	9.88	0.85	9.0	9.38	9.75	10.25	11	

```
naniar::miss_var_summary(SC)
```

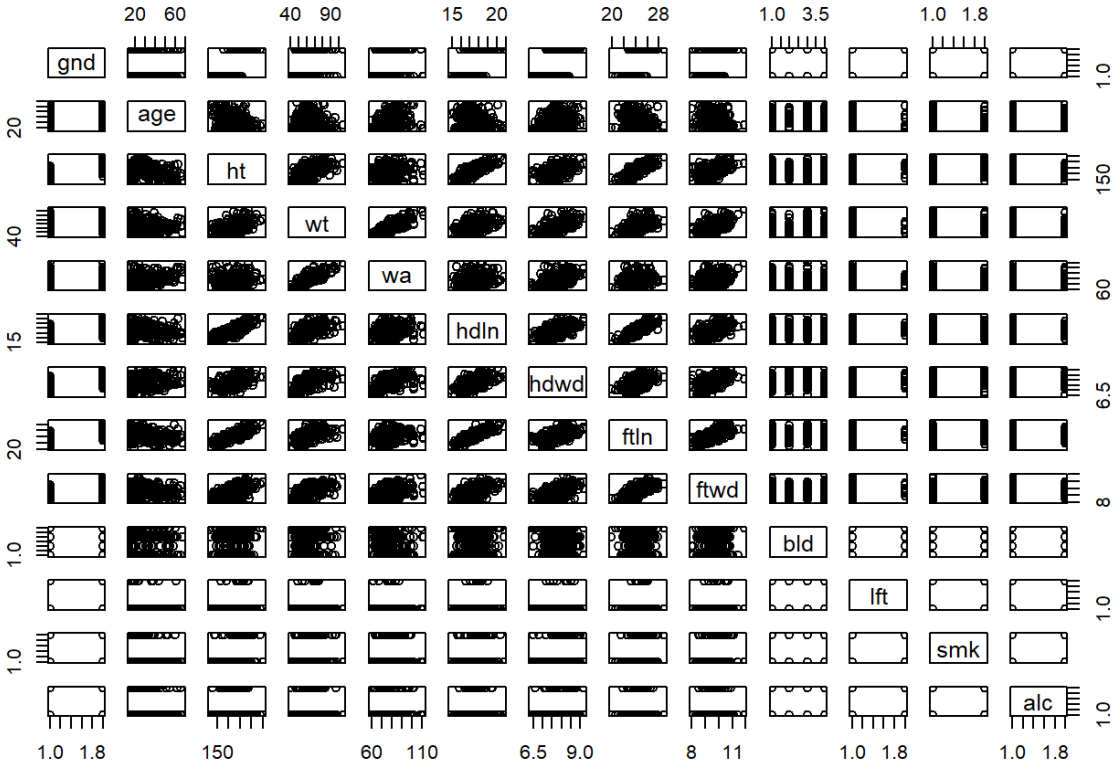
```
## # A tibble: 13 x 3
##    variable n_miss pct_miss
##    <chr>     <int>    <dbl>
##  1 gnd           0        0
##  2 age           0        0
##  3 ht            0        0
##  4 wt            0        0
##  5 wa            0        0
##  6 hdln          0        0
##  7 hdwd          0        0
##  8 ftln          0        0
##  9 ftwd          0        0
## 10 bld           0        0
## 11 lft           0        0
## 12 smk           0        0
## 13 alc           0        0
```

```
DF_nomiss <- na.omit(DF)
head(DF_nomiss)
```

```
##   gnd age    ht   wt   wa hdln hdwd ftln ftwd bld lft smk alc
## 1   M  42 165.0 79.0 96.4 17.9  8.6 24.5  9.6   0   N   N   Y
## 2   M  23 187.5 77.8 76.6 20.4  8.3 28.9 10.6   0   N   Y   N
## 3   F  32 161.7 59.6 83.0 17.5  7.7 23.3 10.4   A   N   N   N
## 4   F  30 162.0 52.0 65.8 16.4  6.6 23.7  8.4   B   N   N   N
## 6   M  26 179.0 76.3 83.8 19.0  8.6 25.7 10.7   B   N   N   N
## 7   M  20 176.6 61.4 70.1 19.8  7.8 26.5  9.6  AB   N   N   N
```

간단 탐색

```
plot(DF_nomiss)
```



- 연속 ~ 연속

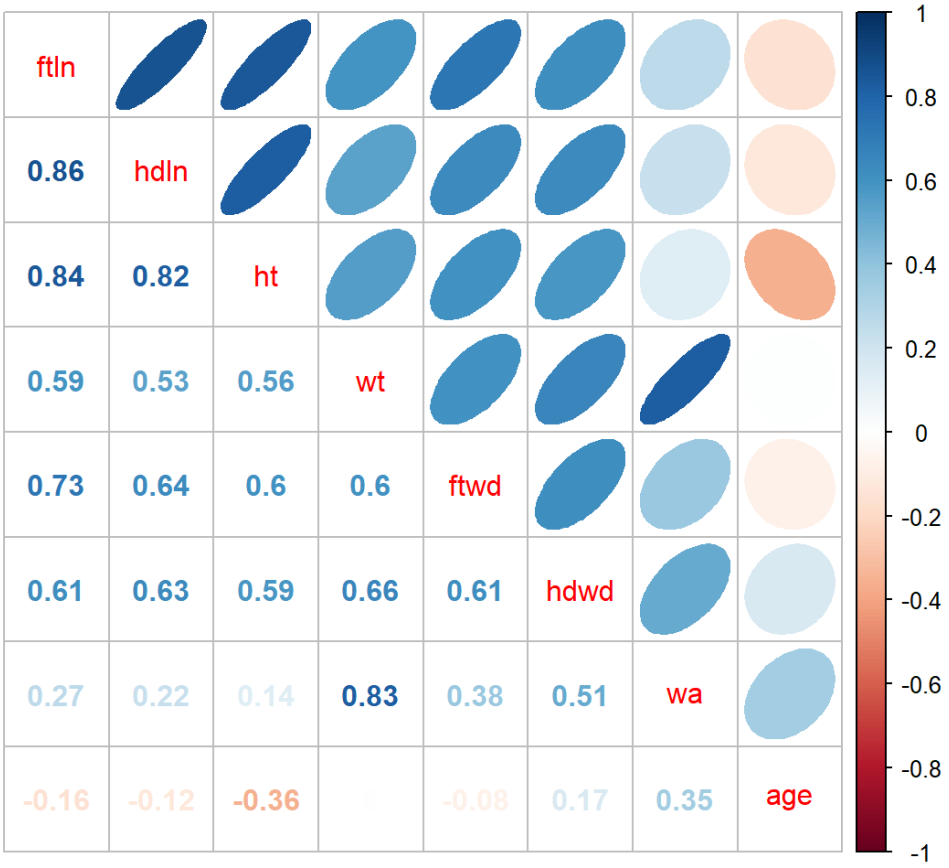
```
R <- cor (DF_nomiss%>% select_if(is.numeric), use='pairwise.complete.obs')
R
```

```
##                age         ht           wt        wa       hdln      hdwd
## age    1.0000000000 -0.3583315 0.0002809797 0.3476877 -0.1224904 0.1689106
## ht    -0.3583314538  1.0000000 0.5591747120 0.1351211  0.8243131 0.5858071
## wt     0.0002809797  0.5591747 1.0000000000 0.8291867  0.5346616 0.6562938
## wa     0.3476876598  0.1351211 0.8291866702 1.0000000  0.2206278 0.5082119
## hdln  -0.1224903692  0.8243131 0.5346615754 0.2206278  1.0000000 0.6348799
## hdwd   0.1689105886  0.5858071 0.6562937725 0.5082119  0.6348799 1.0000000
## ftln  -0.1564673132  0.8406184 0.5916528574 0.2655957  0.8649762 0.6134163
## ftwd  -0.0759035083  0.6025835 0.6004335116 0.3785930  0.6374492 0.6118779
##             ftln        ftwd
## age   -0.1564673 -0.07590351
## ht     0.8406184  0.60258347
## wt     0.5916529  0.60043351
## wa     0.2655957  0.37859296
## hdln   0.8649762  0.63744924
## hdwd   0.6134163  0.61187788
## ftln   1.0000000  0.72924437
## ftwd   0.7292444  1.00000000
```

```
sort(R['ht', ], decreasing=TRUE)
```

```
##         ht       ftln       hdln       ftwd       hdwd         wt         wa
##  1.0000000  0.8406184  0.8243131  0.6025835  0.5858071  0.5591747  0.1351211
##        age
## -0.3583315
```

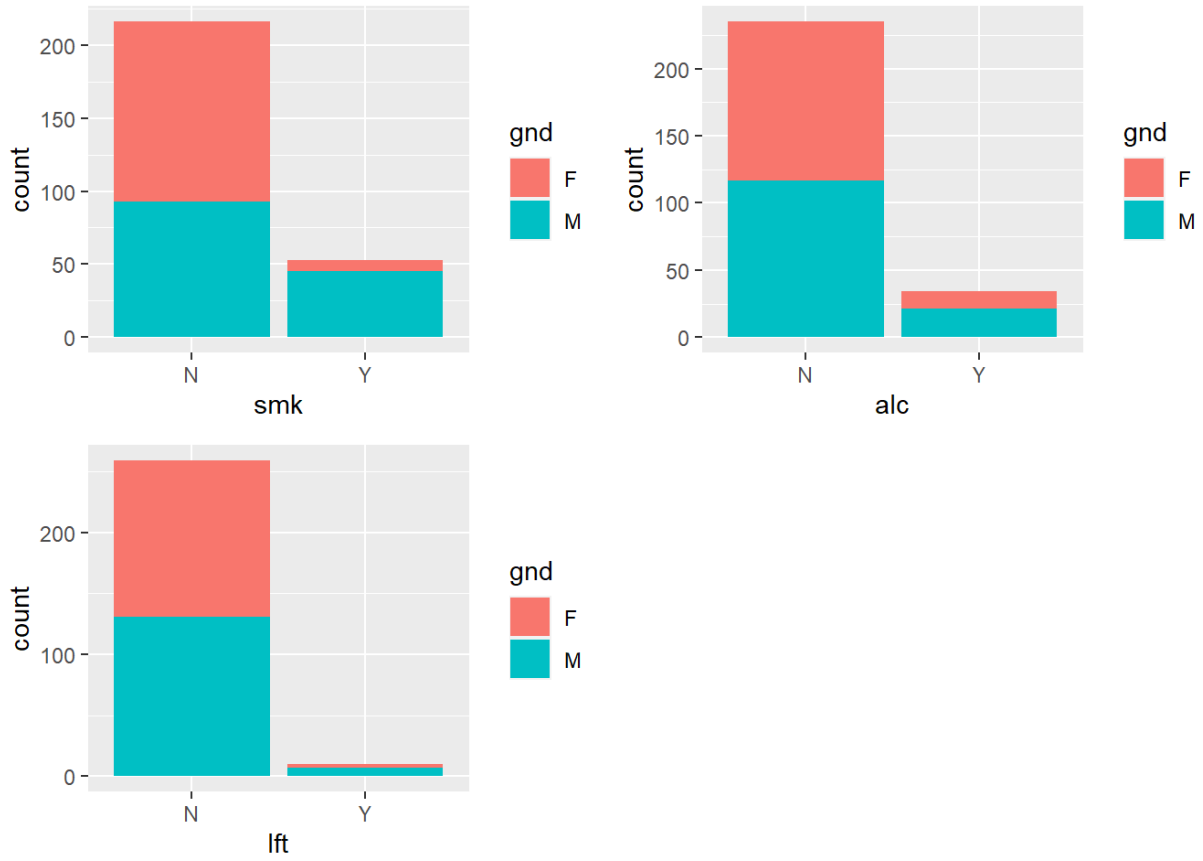
```
corrplot::corrplot.mixed(R, upper='ellipse', order='FPC')
```



- 이산 ~ 이산
  - 성별 흡연비율을 시각화
  - 성별 음주비율을 시각화
  - 성별 왼손잡이비율을 시각화

```
g1 <- ggplot(DF_nomiss, aes(x=smk, fill=gnd)) + geom_bar()
g2 <- ggplot(DF_nomiss, aes(x=alc, fill=gnd)) + geom_bar()
g3 <- ggplot(DF_nomiss, aes(x=lft, fill=gnd)) + geom_bar()

grid.arrange(g1, g2, g3, nrow=2, ncol=2)
```



회귀분석

선형회귀분석

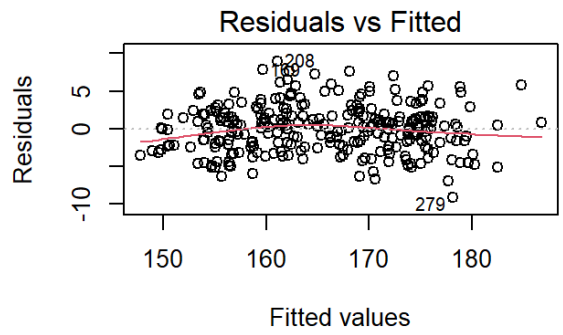
```
Mlm <- lm(ht~., data=DF_nomiss)
summary(Mlm)
```

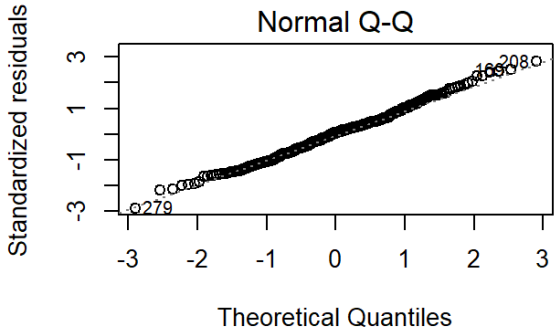
```
##
## Call:
## lm(formula = ht ~ ., data = DF_nomiss)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.1175 -2.2610  0.1498  1.8112  8.9419
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 104.86782    6.51222  16.103  < 2e-16 ***
## gndM          5.31866    0.68905   7.719 2.69e-13 ***
## age          -0.06027    0.02098  -2.873  0.00441 **
## wt            0.39190    0.05044   7.770 1.95e-13 ***
## wa           -0.38732    0.05448  -7.110 1.18e-11 ***
## hdln          2.12427    0.38113   5.574 6.36e-08 ***
## hdwd          0.02702    0.66800   0.040  0.96777
## ftln          1.57590    0.31751   4.963 1.27e-06 ***
## ftwd         -1.08832    0.45006  -2.418  0.01630 *
## bldAB         0.55505    0.68748   0.807  0.42021
## bldB         -0.31038    0.49828  -0.623  0.53390
## bldO         -0.13907    0.56166  -0.248  0.80463
## lftY          0.18510    1.05292   0.176  0.86059
## smkY         -0.32792    0.54048  -0.607  0.54458
## alcY         -0.43018    0.60414  -0.712  0.47708
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.229 on 254 degrees of freedom
## Multiple R-squared:  0.8802, Adjusted R-squared:  0.8736
## F-statistic: 133.3 on 14 and 254 DF,  p-value: < 2.2e-16
```

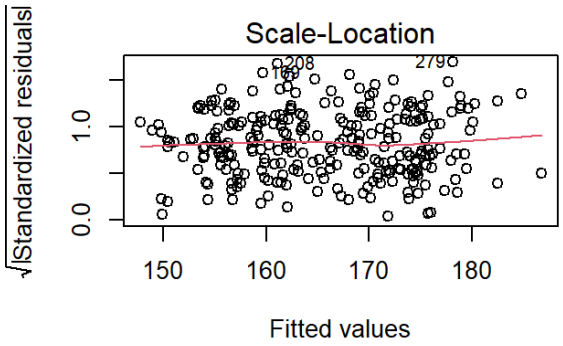
Mlm 식:

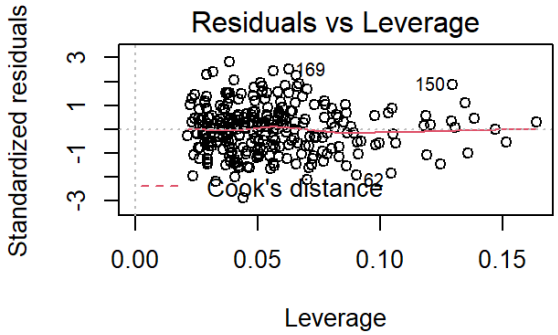
y = 104.86782 + 5.31866gndM - 0.06027age + 0.39190wt -0.38732wa + 2.12427hdwd + 1.57590ftln - 1.08832ftwd + 0.55505bldAB - 0.31038bldB -0.13907bldO + 0.18510lftY - 0.32792smkY - 0.43018alcY

```
par(mfrow=c(2,2))
plot(Mlm)
```










```
predict(Mlm,SC)
```

```
##        1        2        3        4
## 153.7971 166.2341 176.3076 164.3920
```

```
y = SC$ht
pred=predict(Mlm, SC)
resid = y-pred
cbind(y, pred, resid)
```

```
##     y     pred     resid
## 1 154 153.7971  0.202867
## 2 169 166.2341  2.765890
## 3 174 176.3076 -2.307604
## 4 155 164.3920 -9.392002
```

```
MSE = sum(resid^2)/4
MAE = sum(abs(resid))/4

cat("MSE: ",  MSE, "   MAE: ", MAE)
```

```
## MSE:  25.30651    MAE:  3.667091
```

RSQ, MSE, MAE

Rsq: 30.8802, MSE: 25.30651, MAE: 3.667091

변수선택 회귀분석

```
Mstep <- stepAIC(Mlm)
```

```
## Start:  AIC=645.15
## ht ~ gnd + age + wt + wa + hdln + hdwd + ftln + ftwd + bld +
##     lft + smk + alc
##
##        Df Sum of Sq    RSS    AIC
## - bld   3     18.19 2666.0 640.99
## - hdwd  1      0.02 2647.9 643.15
## - lft   1      0.32 2648.2 643.18
## - smk   1      3.84 2651.7 643.54
## - alc   1      5.29 2653.1 643.68
## <none>              2647.8 645.15
## - ftwd  1     60.96 2708.8 649.27
## - age   1     86.06 2733.9 651.75
## - ftln  1    256.81 2904.7 668.05
## - hdln  1    323.84 2971.7 674.18
## - wa    1    526.92 3174.8 691.97
## - gnd   1    621.11 3268.9 699.83
## - wt    1    629.32 3277.2 700.51
##
## Step:  AIC=640.99
## ht ~ gnd + age + wt + wa + hdln + hdwd + ftln + ftwd + lft +
##     smk + alc
##
##        Df Sum of Sq    RSS    AIC
## - hdwd  1      0.03 2666.1 638.99
## - lft   1      0.21 2666.2 639.01
## - smk   1      2.20 2668.2 639.21
## - alc   1      4.95 2671.0 639.49
## <none>              2666.0 640.99
## - ftwd  1     57.01 2723.0 644.68
## - age   1     89.58 2755.6 647.88
## - ftln  1    243.89 2909.9 662.53
## - hdln  1    339.42 3005.4 671.22
## - wa    1    537.46 3203.5 688.39
## - wt    1    637.06 3303.1 696.63
## - gnd   1    644.46 3310.5 697.23
##
## Step:  AIC=638.99
## ht ~ gnd + age + wt + wa + hdln + ftln + ftwd + lft + smk + alc
##
##        Df Sum of Sq    RSS    AIC
## - lft   1      0.22 2666.3 637.01
## - smk   1      2.26 2668.3 637.22
## - alc   1      4.93 2671.0 637.49
## <none>              2666.1 638.99
## - ftwd  1     58.38 2724.4 642.82
## - age   1    101.65 2767.7 647.06
## - ftln  1    248.44 2914.5 660.96
## - hdln  1    360.46 3026.5 671.10
## - wa    1    537.85 3203.9 686.42
## - wt    1    667.87 3333.9 697.13
## - gnd   1    841.73 3507.8 710.80
##
## Step:  AIC=637.01
## ht ~ gnd + age + wt + wa + hdln + ftln + ftwd + smk + alc
##
##        Df Sum of Sq    RSS    AIC
## - smk   1      2.22 2668.5 635.24
## - alc   1      4.99 2671.3 635.52
## <none>              2666.3 637.01
## - ftwd  1     58.23 2724.5 640.83
## - age   1    101.45 2767.7 645.06
## - ftln  1    249.07 2915.4 659.04
```

```
## - hdln  1    360.24 3026.5 669.10
## - wa    1    539.18 3205.5 684.56
## - wt    1    668.33 3334.6 695.18
## - gnd   1    843.16 3509.4 708.93
##
## Step:  AIC=635.24
## ht ~ gnd + age + wt + wa + hdln + ftln + ftwd + alc
##
##        Df Sum of Sq    RSS    AIC
## - alc   1      5.94 2674.4 633.84
## <none>              2668.5 635.24
## - ftwd  1     57.73 2726.2 638.99
## - age   1    106.14 2774.6 643.73
## - ftln  1    247.10 2915.6 657.06
## - hdln  1    364.84 3033.3 667.71
## - wa    1    537.06 3205.6 682.56
## - wt    1    666.61 3335.1 693.22
## - gnd   1    858.20 3526.7 708.25
##
## Step:  AIC=633.84
## ht ~ gnd + age + wt + wa + hdln + ftln + ftwd
##
##        Df Sum of Sq    RSS    AIC
## <none>              2674.4 633.84
## - ftwd  1     57.46 2731.9 637.55
## - age   1    107.84 2782.3 642.47
## - ftln  1    243.15 2917.6 655.24
## - hdln  1    371.49 3045.9 666.82
## - wa    1    542.54 3217.0 681.52
## - wt    1    672.50 3346.9 692.17
## - gnd   1    852.52 3527.0 706.27
```

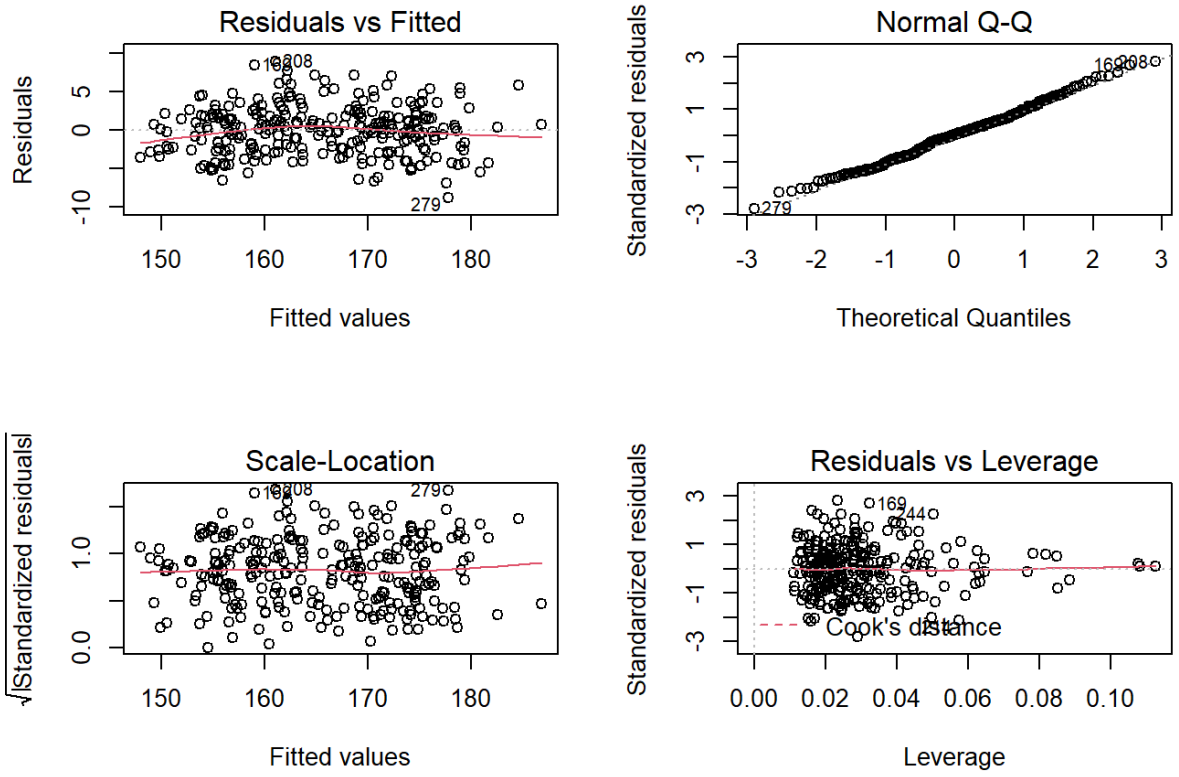
```
summary(Mstep)
```

```
##
## Call:
## lm(formula = ht ~ gnd + age + wt + wa + hdln + ftln + ftwd, data = DF_nomiss)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8275 -2.4071  0.1177  1.8689  8.9437
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.41125    5.70777  18.468  < 2e-16 ***
## gndM          5.27676    0.57851   9.121  < 2e-16 ***
## age          -0.06226    0.01919  -3.244  0.00133 **
## wt            0.39174    0.04836   8.101 2.10e-14 ***
## wa           -0.38821    0.05335  -7.276 4.02e-12 ***
## hdln          2.19044    0.36379   6.021 5.85e-09 ***
## ftln          1.49213    0.30632   4.871 1.93e-06 ***
## ftwd         -1.03522    0.43716  -2.368  0.01861 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.201 on 261 degrees of freedom
## Multiple R-squared:  0.879,  Adjusted R-squared:  0.8758
## F-statistic: 270.9 on 7 and 261 DF,  p-value: < 2.2e-16
```

Mstep 식:

y = 105.41125 + 5.27676gndM -0.06226age + 0.39174wt - 0.38821wa + 2.19044hdln + 1.49213ftln - 1.03522ftwd

```
par(mfrow=c(2,2))
plot(Mstep)
```




```
predict(Mstep, SC)
```

```
##        1        2        3        4
## 153.2895 166.6282 176.7933 163.5287
```

```
y = SC$ht
pred=predict(Mstep, SC)
resid = y-pred
cbind(y, pred, resid)
```

```
##     y     pred      resid
## 1 154 153.2895  0.7105283
## 2 169 166.6282  2.3718285
## 3 174 176.7933 -2.7933482
## 4 155 163.5287 -8.5287394
```

```
MSE = sum(resid^2)/4
MAE = sum(abs(resid))/4

cat("MSE: ",  MSE, "   MAE: ", MAE)
```

```
## MSE:  21.66815    MAE:  3.601111
```

RSQ, MSE, MAE

Rsq: 0.879, MSE: 21.66815, MAE: 3.601111