```
load(url("http://ph.emu.ee/~ktanel/DK_0016/students.RData"))
```

1. Calculate the students' body mass index (as a new variable) and study its distribution separately for men and women. Does the body mass index follow the normal distribution?

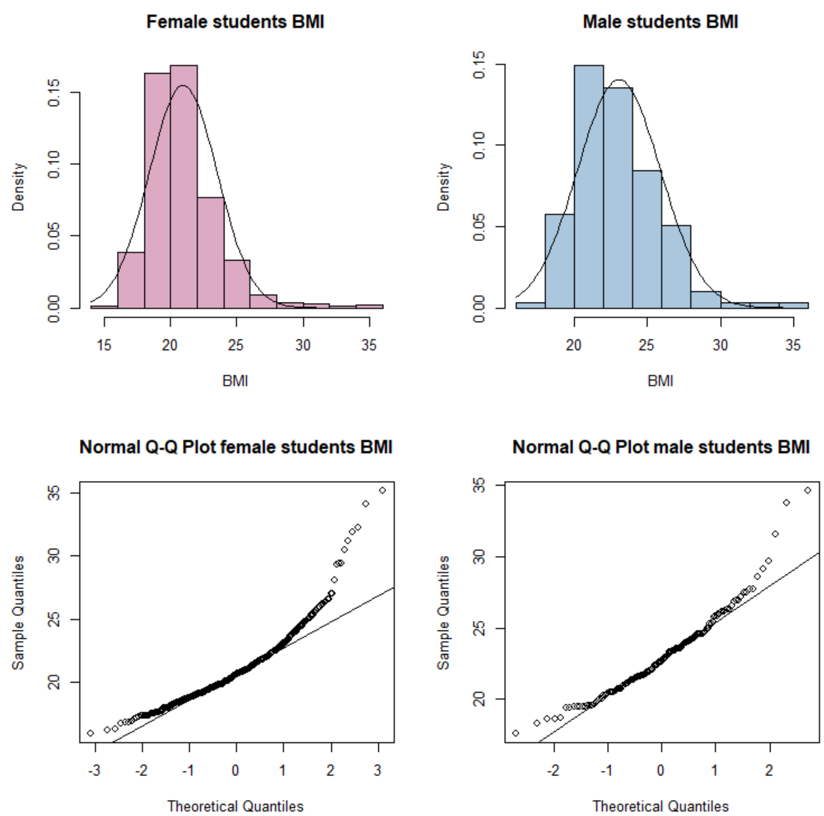
Code:

```
#----------------------------------------------------------------
#-----1-----
# Calculate the students' body mass index (as a new variable)

students$bmi <- students$weight / ((students$height / 100)^2)

# and study its distribution separately for men and women
# Does the body mass index follow the normal distribution?

summary(students$bmi)
windows(8, 8)
par(mfrow = c(2, 2))
    hist(students$bmi[students$gender == "1"],
        main = "Female students BMI",
        xlab = "BMI",
        col = "#DDAAC4",
        freq = FALSE
    )
        curve(
            dnorm(
                x, mean = mean(students$bmi[students$gender == "1"], na.rm = T),
                sd = sd(students$bmi[students$gender == "1"], na.rm = T)
            ),
            add = T
        )

     hist(students$bmi[students$gender == "2"],
        main = "Male students BMI",
        xlab = "BMI",
        col = "#AAC7DD",
        freq = FALSE
    )
```

```
        curve(
            dnorm(
                x, mean = mean(students$bmi[students$gender == "2"], na.rm = T),
                sd = sd(students$bmi[students$gender == "2"], na.rm = T)
            ),
            add = T
        )
    qqnorm(
        students$bmi[students$gender == "1"],
        main = "Normal Q-Q Plot female students BMI"
    )
        qqline(students$bmi[students$gender == "1"])
    qqnorm(
        students$bmi[students$gender == "2"],
        main = "Normal Q-Q Plot male students BMI"
    )
        qqline(students$bmi[students$gender == "2"])
```

Outputs:

```
r$> summary(students$bmi)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  15.92   19.49   20.94   21.42   22.83   35.16       4 
```



Answer:

Body Mass Index does not follow a normal distribution, either for female or male students.

2. What is the average (± standard deviation) body mass index of men and women? Is the difference statistically significant? Make the decision both based on the 95% confidence interval of means' difference and based on the p-value.

Code:

```
#-----2-----
# What is the average (± standard deviation) body mass index of men and women?

mean(students$bmi [students$gender == 1], na.rm = TRUE)
sd(students$bmi[students$gender == 1], na.rm = TRUE)
mean(students$bmi [students$gender == 2], na.rm = TRUE)
sd(students$bmi[students$gender == 2], na.rm = TRUE)

# Is the difference statistically significant?
# Make the decision both based on the 95% confidence interval
# of means' difference and based on the p-value.

t.test(bmi ~ gender, data = students)
```

Outputs:

```
r$> mean(students$bmi [students$gender == 1], na.rm = TRUE)
[1] 20.95111

r$> sd(students$bmi [students$gender == 1], na.rm = TRUE)
[1] 2.573081

r$> mean(students$bmi [students$gender == 2], na.rm = TRUE)
[1] 23.03513

r$> sd(students$bmi [students$gender == 2], na.rm = TRUE)
[1] 2.843128

r$> t.test(bmi ~ gender, data = students)

        Welch Two Sample t-test

data:  bmi by gender
t = -8.0125, df = 221.86, p-value = 6.342e-14
alternative hypothesis: true difference in means between group 1 and group 2
is not equal to 0
95 percent confidence interval:
 -2.596593 -1.571440
sample estimates:
mean in group 1 mean in group 2
       20.95111        23.03513
```

<u>Answer:</u>

$$\bar{x}(\,BMI_{women}) = \; 20.95 \pm 2.57$$

$$\bar{x}(\,BMI_{men}) = \; 23.04 \pm 2.84$$

```
Welch Two Sample t-test
```

$H_o\!: \bar{x}(\,BMI_{women}) = \bar{x}(\,BMI_{men})$

$H_1\!: \bar{x}(\,BMI_{women}) \neq \bar{x}(\,BMI_{men})$

$$p - value = \; 6.342e^{-14} \leq 0.05$$

Women's BMI is statistically significantly different from men's BMI.

3. Create a new variable 'sport01' with value zero, if student does not practice sport (sport=1), and value one, if student practices sport (sport>1). You can use the following command: `sport01 <- factor(sport, levels=1:5, labels=c(0,1,1,1,1))`. How big is the percentage of students practicing sport (with 95% confidence interval)? Is the percentage of students practicing sport statistically significant from 75%?

Code:

```
#-----3-----
# Create a new variable 'sport01' with value zero,
# if student does not practice sport (sport=1), and value one,
# if student practices sport (sport>1).
# You can use the following command:

students$sport01 <- factor(
    students$sport,
    levels = 1:5,
    labels = c(0, 1, 1, 1, 1)
)

# How big is the percentage of students practicing sport
# (with 95% confidence interval)? Is the percentage of
# students practicing sport statistically significant from 75%?

table(students$sport01)
prop.test(
    c(124, 535),
    c(124 + 535, 124 + 535)
)
prop.test(535, 660, p = 0.75)
```

Outputs:

```
r$> table(students$sport01)

  0   1 
124 535 

r$> prop.test(
        c(124, 535),
        c(124 + 535, 124 + 535)
    )

        2-sample test for equality of proportions with continuity correction

data:  c(124, 535) out of c(124 + 535, 124 + 535)
X-squared = 510.17, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
```

```
95 percent confidence interval:
 -0.6673907 -0.5799538
sample estimates:
   prop 1    prop 2
0.1881639 0.8118361

r$> prop.test(535, 660, p = 0.75)

        1-sample proportions test with continuity correction

data:  535 out of 660, null probability 0.75
X-squared = 12.608, df = 1, p-value = 0.0003841
alternative hypothesis: true p is not equal to 0.75
95 percent confidence interval:
 0.7781501 0.8393766
sample estimates:
        p
0.8106061
```

Answer:

`2-sample test for equality of proportions with continuity correction`

H_o: proportion of students practicing sport = proportion of students not practicing sports

H_1: proportion of students practicing sport ≠ proportion of students not practicing sports

$$\%_{sporting\ students} = 0.81$$

`1-sample proportions test with continuity correction (0.75)`

H_o: proportion of students practicing sport = 0.75

H_1: proportion of students practicing sport ≠ 0.75

$$p - value = \ 0.0003841 \leq 0.05$$

The proportion of students practicing sport is statistically significantly different from 75%.

4. What about the percentage of sporting students among men and women? Are these proportions statistically significantly different?

Code:

```
#-----4-----
# What about the percentage of sporting students among men and women?
# Are these proportions statistically significantly different?

table(
    students$gender,
    students$sport01 == "1"
)
prop.test(
    c(408, 127),
    c(408 + 103, 127 + 21)
)
```

Outputs:

```
r$> table(
        students$gender,
        students$sport01 == "1"
    )

    FALSE TRUE
  1   103  408
  2    21  127

r$> prop.test(
        c(408, 127),
        c(408 + 103, 127 + 21)
    )

        2-sample test for equality of proportions with continuity correction

data:  c(408, 127) out of c(408 + 103, 127 + 21)
X-squared = 2.2988, df = 1, p-value = 0.1295
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1301379  0.0107906
sample estimates:
   prop 1    prop 2
0.7984344 0.8581081
```

Answer:

`2-sample test for equality of proportions with continuity correction`

H_o: proportion of women practicing sport = proportion of men practicing sport

H_1: proportion of women practicing sport ≠ proportion of men practicing sport

$$p - value = 0.1295 > 0.05$$

The proportion of women practicing sport is not statistically significantly different from the proportion of men practicing sport.

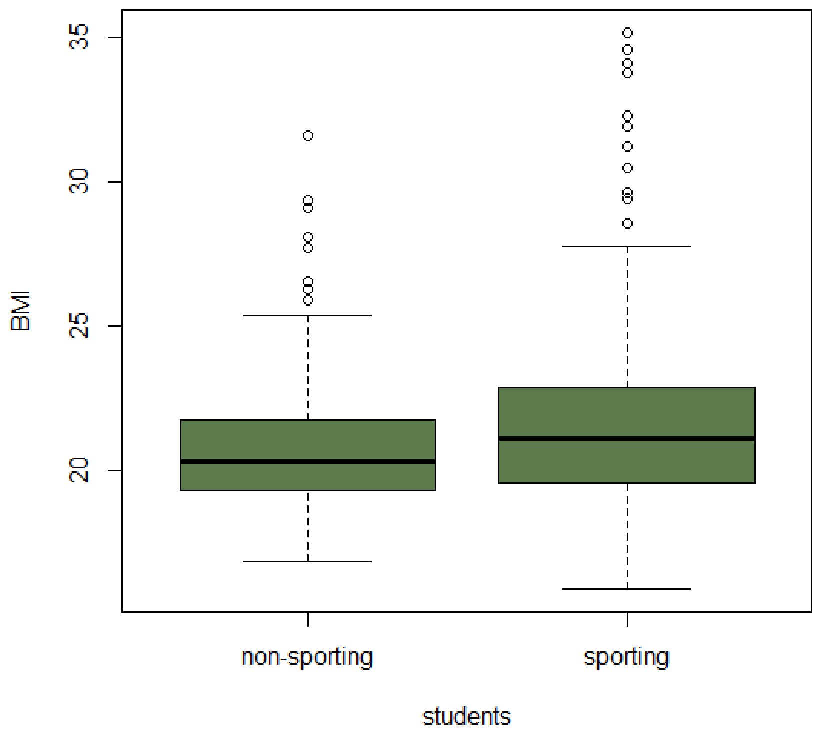
5. Are the body mass index and sporting associated? Test the statistical significance of body mass index difference between sporting and non-sporting students. Visualize the result.

Code:

```
#-----5-----
# Are the body mass index and sporting associated?
# Test the statistical significance of body mass index difference
# between sporting and non-sporting students.

t.test(
    students$bmi[students$sport01 == 1],
    students$bmi[students$sport01 == 0]
)

# Visualize the result

windows(6, 6)
boxplot(
    bmi ~ sport01,
    data = students,
    names = c("non-sporting", "sporting"),
    xlab = "students",
    ylab = "BMI",
    col = "#5d7c4d"
)
```

Outputs:

```
r$> t.test(
        students$bmi[students$sport01==1],
        students$bmi[students$sport01==0]
    )

        Welch Two Sample t-test

data:  students$bmi[students$sport01 == 1] and students$bmi[students$sport01
== 0]
t = 1.8352, df = 188.67, p-value = 0.06805
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.03726451  1.03263787
sample estimates:
mean of x mean of y
 21.51536  21.01767
```



Answer:

`Welch Two Sample t-test`

$H_o: \bar{x}(\,BMI_{no\ sporting}) = \bar{x}(\,BMI_{sporting})$

$H_1: \bar{x}(\,BMI_{no\ sporting}) \neq \bar{x}(\,BMI_{sporting})$

$$p - value = \ 0.06805 > 0.05$$

BMI of non-sporting students is not statistically significantly different from BMI of sporting students.

6. Are the sporting and health associated? Test, visualize and comment the result.

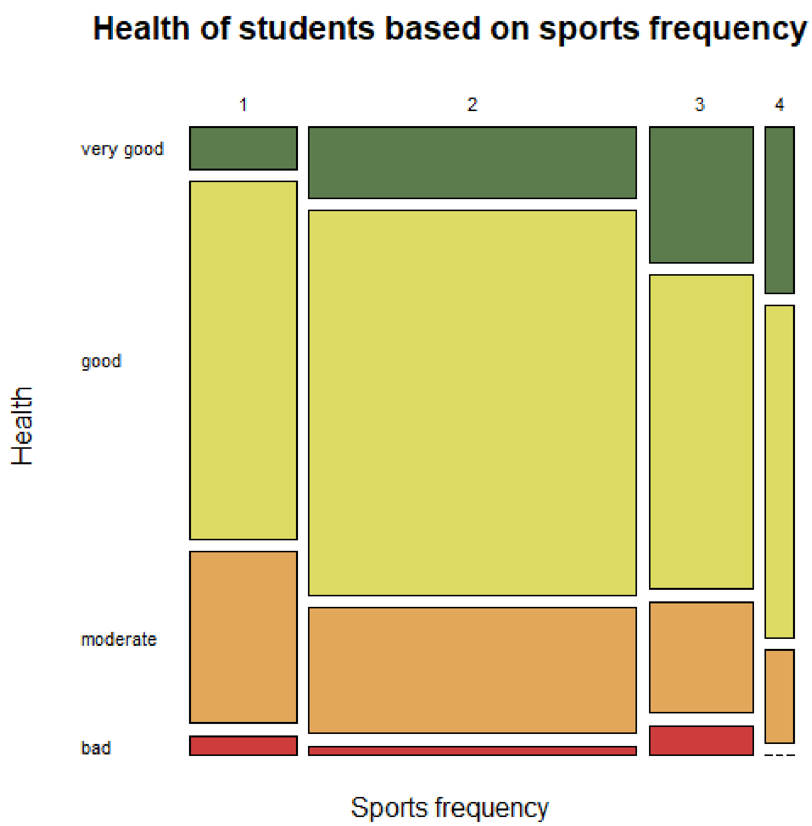
<u>Code:</u>

```
#-----6-----
# Are the sporting and health associated?
# Test, visualize and comment the result.

prop.table(
    table(
        students$sport,
        students$health
    ),
    1
)
chisq.test(
    table(
        students$sport,
        students$health
    ),
    simulate = TRUE
)

windows(6, 6)
plot(
    table(students$sport, students$health),
    las = 1,
    main = "Health of students based on sports frequency",
    xlab = "Sports frequency",
    ylab = "Health",
    col = c("#5d7c4d", "#dddb62", "#e2a759", "#ce3c3c")
)
```

<u>Outputs:</u>

```
r$> prop.table(
        table(
            students$sport,
            students$health
        ),
        1
    )

     very good       good   moderate        bad
  1 0.07258065 0.60483871 0.29032258 0.03225806
  2 0.12073491 0.65091864 0.21259843 0.01574803
  3 0.22950820 0.53278689 0.18852459 0.04918033
  4 0.28125000 0.56250000 0.15625000 0.00000000

r$> chisq.test(
```

```
        table(
            students$sport,
            students$health
        ),
        simulate = TRUE
    )

        Pearson's Chi-squared test with simulated p-value (based on 2000
replicates)

data:  table(students$sport, students$health)
X-squared = 28.305, df = NA, p-value = 0.003498
```



Answer:

`Pearson's Chi-squared test with simulated p-value (based on 2000 replicates`

H_o: sporting and health are not associated

H_1: sporting and health are associated

$$p - value = 0.003498 \leq 0.05$$

Sporting frequency and health status of students are associated.

7.

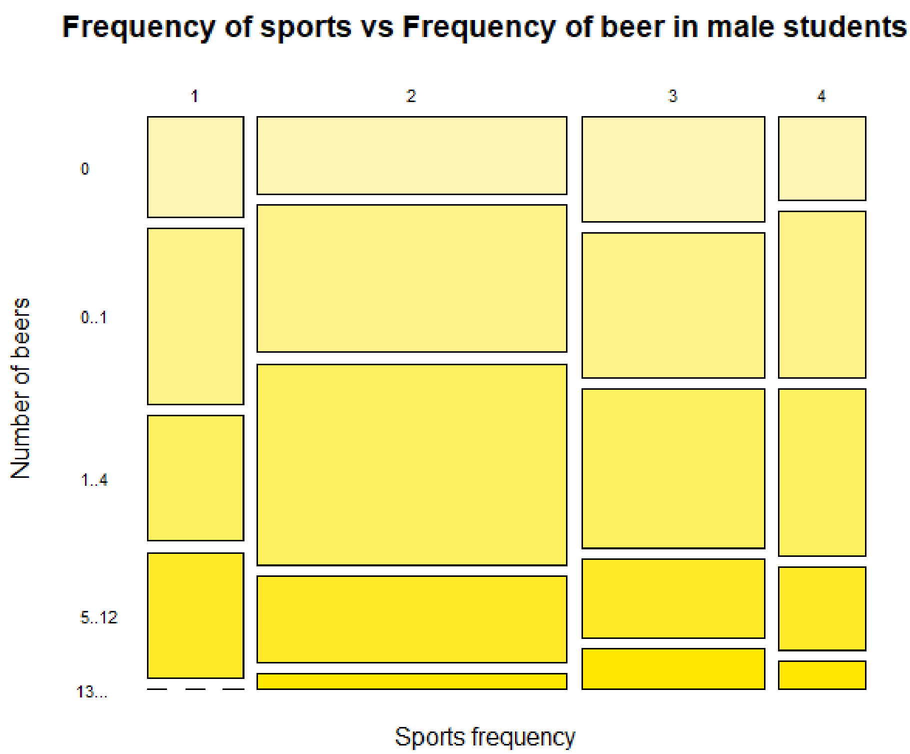
Code:

```
#-----7-----
# Are the beer consumption and sporting of male students associated?

prop.table(
    table(
        students$sport[students$gender == 2],
        students$beer[students$gender == 2]
    ),
    1
 )
chisq.test(
    table(
        students$sport[students$gender == 2],
        students$beer[students$gender == 2]),
    simulate = TRUE
)

windows(7, 6)
plot(
    table(
        students$sport[students$gender == 2],
        students$beer[students$gender == 2]),
    las = 1,
    main = "Frequency of sports vs Frequency of beer in male students",
    xlab = "Sports frequency",
    ylab = "Number of beers",
    col = c("#FCF5B3", "#FFF489", "#FFF060", "#FFEB29", "#FFE700")
)
```

Outputs:

```
r$> prop.table(
        table(
            students$sport[students$gender == 2],
            students$beer[students$gender == 2]),
            1
        )

              0       0..1       1..4      5..12      13...
  1 0.19047619 0.33333333 0.23809524 0.23809524 0.00000000
  2 0.14705882 0.27941176 0.38235294 0.16176471 0.02941176
  3 0.20000000 0.27500000 0.30000000 0.15000000 0.07500000
  4 0.15789474 0.31578947 0.31578947 0.15789474 0.05263158

r$> chisq.test(
        table(
```

```
            students$sport[students$gender == 2],
            students$beer[students$gender == 2]),
            simulate = TRUE
        )

        Pearson's Chi-squared test with simulated p-value (based on 2000
replicates)

data:  table(students$sport[students$gender == 2],
students$beer[students$gender ==     2])
X-squared = 4.9605, df = NA, p-value = 0.966
```



Answer:

`Pearson's Chi-squared test with simulated p-value (based on 2000 replicates`

H_o: beer consumption and sporting are not associated

H_1: beer consumption and sporting are associated

$$p - value = 0.966 > 0.05$$

The beer consumption and sporting in male students are not associated.

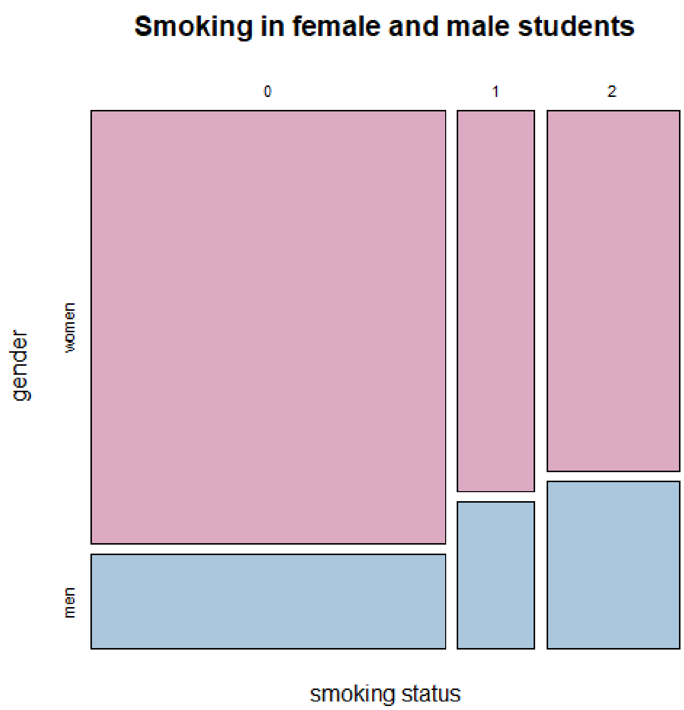
8. Create a new variable ’smoke012’ with value zero, if the student does not smoke (smoking=1), one, if the student no longer smokes but has smoked (smoking=2), and two, if the student smokes (smoking>2). Does the smoking depend on gender? Visualize the result.

Code:

```
#-----8-----
# Create a new variable ’smoke012’ with value zero,
# if the student does not smoke (smoking=1),
# one, if the student no longer smokes but has smoked (smoking=2),
# and two, if the student
# smokes (smoking>2). Does the smoking depend on gender?
# Visualize the result.

students$smoking <- factor(students$smoking)
levels(students$smoking)
students$smoke012 <- factor(
    students$smoking,
    levels = c(1, 2, 3, 4, 5, 6, 7),
    labels = c("0", "1", "2", "2", "2", "2", "2")
)
prop.table(
    table(students$gender, students$smoke012),
    1
)
chisq.test(
    table(students$gender, students$smoke012),
    simulate = TRUE
)
students$gendernames <- factor(
    students$gender,
    levels = 1:2,
    labels = c("women", "men")
)
windows(6, 6)
plot(
    table(students$smoke012, students$gendernames),
    main = "Smoking in female and male students",
    xlab = "smoking status",
    ylab = "gender",
    col = c("#DDAAC4", "#AAC7DD")
)
```

Outputs:

```
r$> prop.table(
        table(students$gender, students$smoke012),
        1
    )

            0         1         2
  1 0.6660156 0.1269531 0.2070312
  2 0.5000000 0.1689189 0.3310811

r$> chisq.test(
        table(students$gender, students$smoke012),
        simulate = TRUE
    )

        Pearson's Chi-squared test with simulated p-value (based on 2000
replicates)

data:  table(students$gender, students$smoke012)
X-squared = 14.038, df = NA, p-value = 0.001999
```



Answer:

`Pearson's Chi-squared test with simulated p-value (based on 2000 replicates`

H_o: smoking and gender are not associated

H_1: smoking and gender are associated

$$p - value = 0.001999 \leq 0.05$$

Smoking depends on gender.

9. Are student body mass index and systolic and diastolic blood pressure related? Whether these relationships are different between men and women (no statistical significance testing of difference is required, but you can visualize the results)?

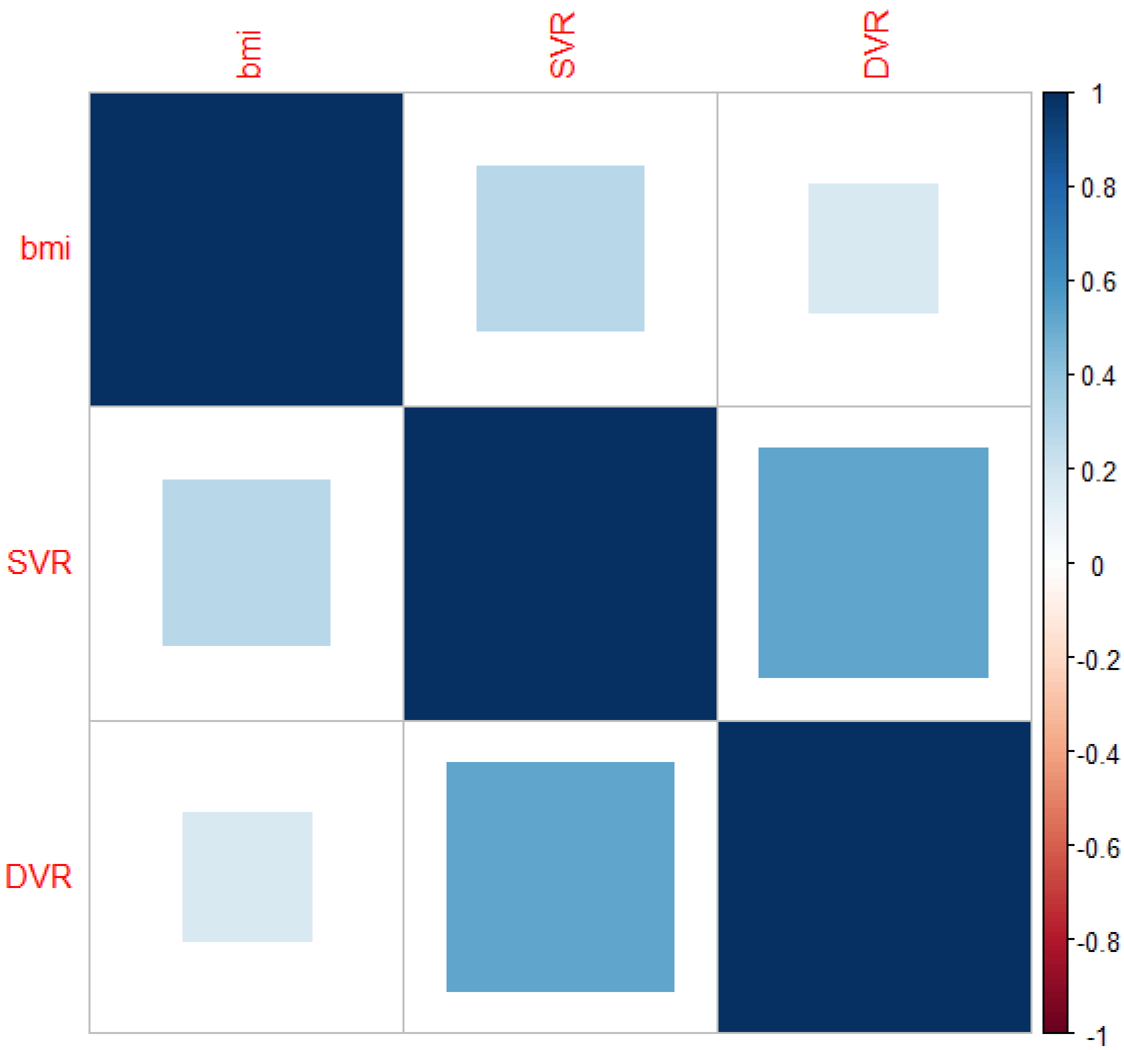
<u>Code:</u>

```
#-----9-----
# Are student body mass index and systolic and diastolic blood pressure related?
# Whether these relationships are different between men and women
# (no statistical significance testing of
# difference is required, but you can visualize the results)?
install.packages("corrplot")
library("corrplot")

# all students
cor(
    x = students [, c("bmi", "SVR", "DVR")],
    use = "complete.obs"
)
correlations <- cor(
    x = students [, c("bmi", "SVR", "DVR")],
    use = "complete.obs"
)
windows(6, 6)
corrplot(
    correlations,
    method = "square",
)

# female students
students_f <- students[students$gender == "1", ]
cor(
    x = students_f [, c("bmi", "SVR", "DVR")],
    use = "complete.obs"
)
correlations_f <- cor(
    x = students_f [, c("bmi", "SVR", "DVR")],
    use = "complete.obs"
)
windows(6, 6)
corrplot(
    correlations_f,
    method = "square",
)

# male students
students_m <- students[students$gender == "2", ]
cor(
    x = students_m [, c("bmi", "SVR", "DVR")],
    use = "complete.obs"
)
```

<u>Outputs:</u>

All students:

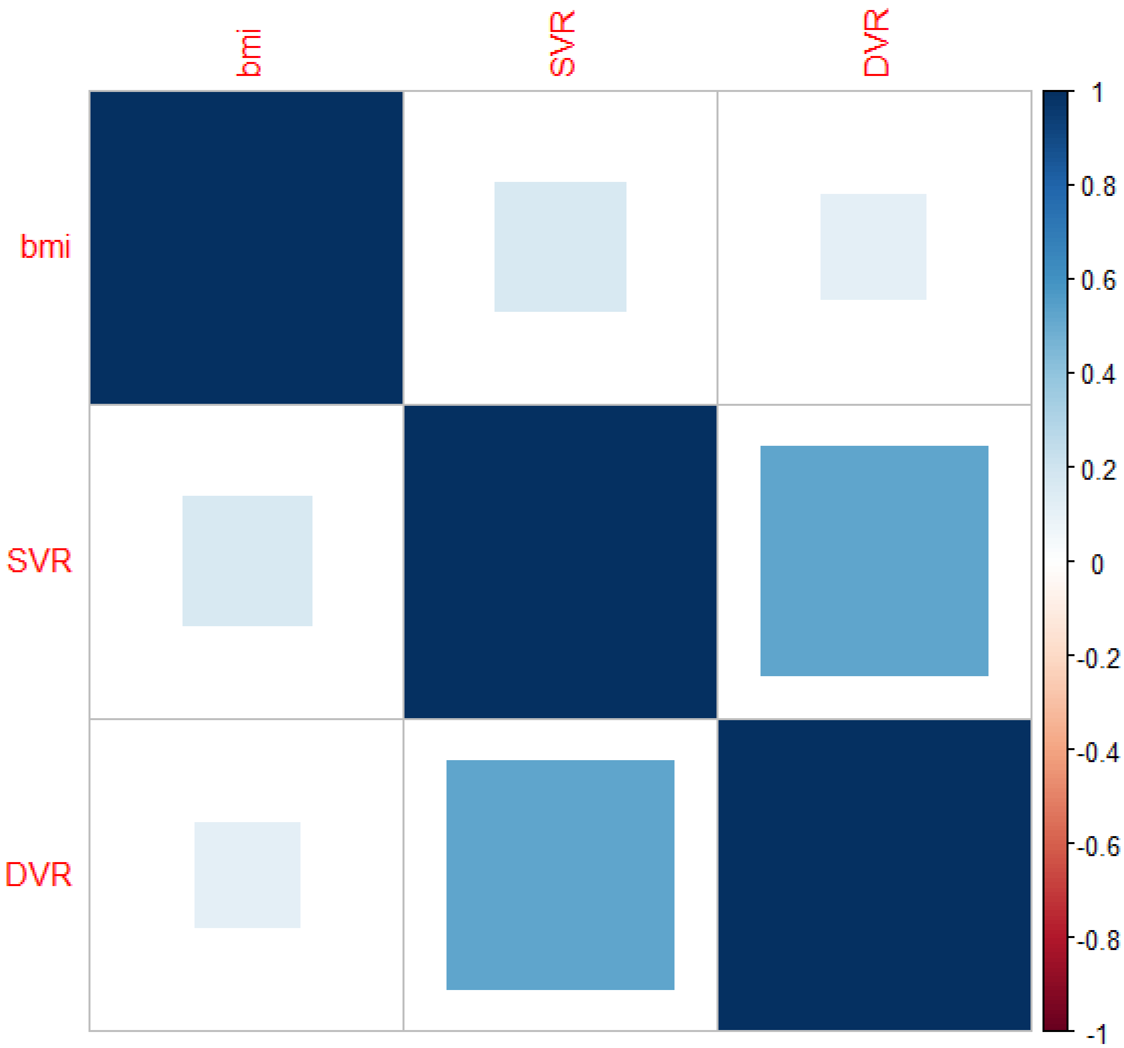
```
r$> cor(
        x = students [, c("bmi", "SVR", "DVR")],
        use = "complete.obs"
    )
          bmi       SVR       DVR
bmi 1.0000000 0.2758059 0.1663297
SVR 0.2758059 1.0000000 0.5269027
DVR 0.1663297 0.5269027 1.0000000
```



Female students:

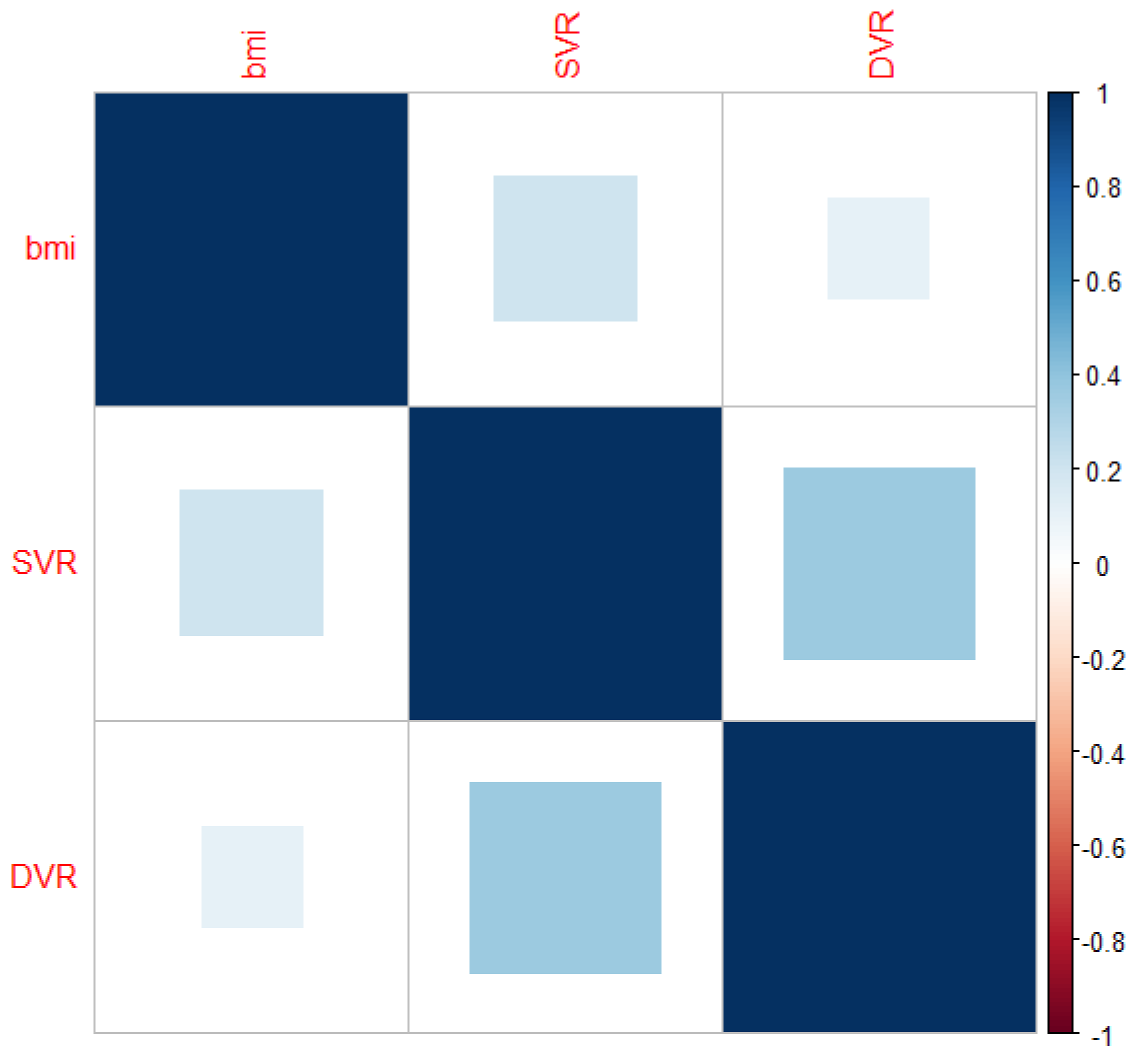
```
r$> cor(
        x = students_f [, c("bmi", "SVR", "DVR")],
        use = "complete.obs"
    )
          bmi       SVR       DVR
bmi 1.0000000 0.1668648 0.1100011
```

```
SVR 0.1668648 1.0000000 0.5231119
DVR 0.1100011 0.5231119 1.0000000
```



Male students:

```
r$> cor(
        x = students_m [, c("bmi", "SVR", "DVR")],
        use = "complete.obs"
    )
          bmi       SVR       DVR
bmi 1.0000000 0.2092348 0.1019517
SVR 0.2092348 1.0000000 0.3664230
DVR 0.1019517 0.3664230 1.0000000
```



Answer:

There’s an intermediate correlation between SVR and DVR of all students ($|r| = 0.53$), only male students ($|r| = 0.52$) and only female students ($|r| = 0.37$) because:

$$0.3 < |r| < 0.7$$

There’s weak correlation for all other combinations because:

$$|r| \leq 0.3$$

10. Predict students’ systolic blood pressure based on gender and body mass index. Illustrate the result of the modeling. What is the expected systolic blood pressure of male and female students with a body mass index of 20 points? But with body mass index of 25 points? Is it necessary to consider also the gender by body mass index interaction? But sporting and/or smoking?

Code:

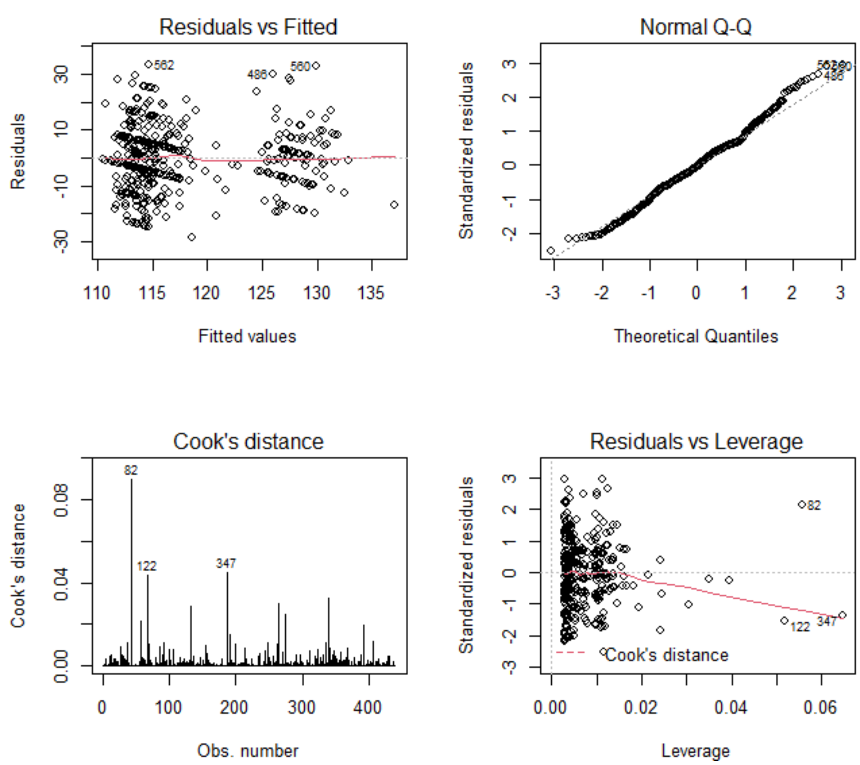
```
#--------------------------------------------------------------
#-----10-----
# Predict students’ systolic blood pressure based on gender and body mass index.
# Illustrate the result of the modeling.

model1 <- lm(SVR ~ gender + bmi, data = students)
summary(model1)

windows()
par(mfrow = c(2, 2))
    plot(model1, 1)
    plot(model1, 2)
    plot(model1, 4)
    plot(model1, 5)


# What is the expected systolic blood pressure of male and female students
# with a body mass index of 20 points? But with body mass index of 25 points?

pred_svr_bmi20_f <- predict(
    model1,
    data.frame(bmi = 20, gender = 1)
)

pred_svr_bmi20_m <- predict(
    model1,
    data.frame(bmi = 20, gender = 2)
)

pred_svr_bmi25_f <- predict(
    model1,
    data.frame(bmi = 25, gender = 1)
)

pred_svr_bmi25_m <- predict(
    model1,
    data.frame(bmi = 25, gender = 2)
)
```

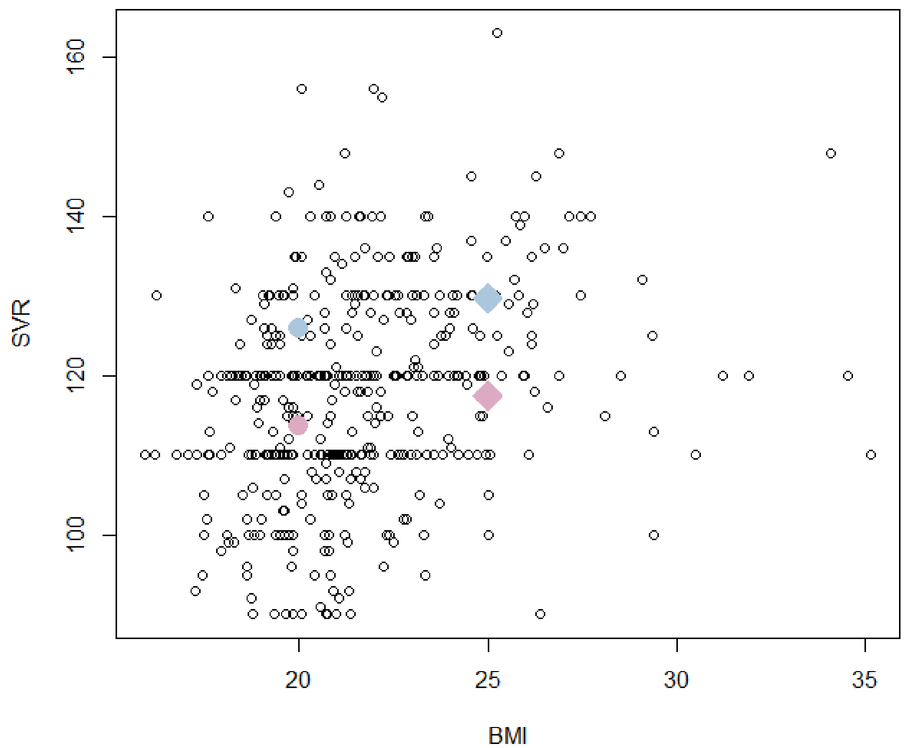
```
windows()
plot(
    SVR ~ bmi,
    data = students,
    xlab = "BMI"
)
    points(
        x = 20,
        y = pred_svr_bmi20_f,
        col = "#DDAAC4",
        pch = 16,
        cex = 2
    )
    points(
        x = 20,
        y = pred_svr_bmi20_m,
        col = "#AAC7DD",
        pch = 16,
        cex = 2
    )
    points(
        x = 25,
        y = pred_svr_bmi25_f,
        col = "#DDAAC4",
        pch = 18,
        cex = 3
    )
    points(
        x = 25,
        y = pred_svr_bmi25_m,
        col = "#AAC7DD",
        pch = 18,
        cex = 3
    )


# Is it necessary to  consider also the gender by body mass index interaction?

model2 <- lm(SVR ~ gender * bmi, data = students)
summary(model2)

anova(model1, model2)
sqrt(56185 / 437) # model 1
sqrt(56176 / 436) # model 2
```

```
# But sporting and/or smoking?

model3 <- lm(SVR ~ gender + bmi + sport + smoking, data = students)
summary(model3)

step_model3 <- step(model3, direction = "both")
```

Outputs:

```
r$> model1 <- lm(SVR ~ gender + bmi, data = students)
    summary(model1)

Call:
lm(formula = SVR ~ gender + bmi, data = students)

Residuals:
    Min      1Q  Median      3Q     Max
-28.550  -6.924  -0.353   6.801  33.398

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  86.0853     4.2486  20.262  < 2e-16 ***
gender       12.2597     1.3289   9.225  < 2e-16 ***
bmi           0.7662     0.2019   3.796 0.000168 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.34 on 437 degrees of freedom
  (220 observations deleted due to missingness)
Multiple R-squared:  0.2278,	Adjusted R-squared:  0.2243
F-statistic: 64.46 on 2 and 437 DF,  p-value: < 2.2e-16
```






```
r$> model2 <- lm(SVR ~ gender * bmi, data = students)

r$> summary(model2)

Call:
lm(formula = SVR ~ gender * bmi, data = students)

Residuals:
    Min      1Q  Median      3Q     Max
-28.419  -6.853  -0.250   6.741  33.398

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.0898    13.8477   6.434 3.29e-10 ***
gender        9.7909    10.9104   0.897    0.370
bmi           0.6317     0.6237   1.013    0.312
gender:bmi    0.1093     0.4792   0.228    0.820
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.35 on 436 degrees of freedom
  (220 observations deleted due to missingness)
Multiple R-squared:  0.2279,    Adjusted R-squared:  0.2226
F-statistic:  42.9 on 3 and 436 DF,  p-value: < 2.2e-16


r$> anova(model1, model2)
Analysis of Variance Table

Model 1: SVR ~ gender + bmi
Model 2: SVR ~ gender * bmi
```

```
  Res.Df   RSS Df Sum of Sq     F Pr(>F)
1    437 56185
2    436 56178  1    6.6972 0.052 0.8198

r$> sqrt(56185 / 437) # model 1
[1] 11.33886

r$> sqrt(56176 / 436) # model 2
[1] 11.35095

r$> model3 <- lm(SVR ~ gender + bmi + sport + smoking, data = students)

r$> summary(model3)

Call:
lm(formula = SVR ~ gender + bmi + sport + smoking, data = students)

Residuals:
    Min      1Q  Median      3Q     Max
-25.890  -7.131   0.043   6.677  33.715

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 87.13673    4.47168  19.486  < 2e-16 ***
gender      11.61004    1.38856   8.361 8.81e-16 ***
bmi          0.72382    0.21053   3.438 0.000643 ***
sport2       0.09001    1.43262   0.063 0.949932
sport3       1.16820    1.79634   0.650 0.515834
sport4       2.74991    2.72998   1.007 0.314361
smoking2     1.43651    1.66103   0.865 0.387619
smoking3    -2.03390    2.03204  -1.001 0.317435
smoking4     0.56077    2.98807   0.188 0.851225
smoking5     1.52638    2.44083   0.625 0.532072
smoking6     2.41649    2.55844   0.945 0.345438
smoking7     0.77545    8.19664   0.095 0.924673
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.4 on 428 degrees of freedom
  (220 observations deleted due to missingness)
Multiple R-squared:  0.2357,    Adjusted R-squared:  0.2161
F-statistic:    12 on 11 and 428 DF,  p-value: < 2.2e-16


r$> step_model3 <- step(model3, direction = "both")
Start:  AIC=2153.29
SVR ~ gender + bmi + sport + smoking

          Df Sum of Sq   RSS    AIC
- smoking  6     416.6 56024 2144.6
- sport    3     201.5 55809 2148.9
<none>                 55607 2153.3
- bmi      1    1535.7 57143 2163.3
- gender   1    9082.9 64690 2217.9

Step:  AIC=2144.57
SVR ~ gender + bmi + sport
```

```
          Df Sum of Sq   RSS    AIC
- sport    3     161.0 56185 2139.8
<none>                 56024 2144.6
+ smoking  6     416.6 55607 2153.3
- bmi      1    1716.5 57740 2155.8
- gender   1   10326.7 66350 2217.0

Step:  AIC=2139.83
SVR ~ gender + bmi

          Df Sum of Sq   RSS    AIC
<none>                 56185 2139.8
+ sport    3     161.0 56024 2144.6
+ smoking  6     376.1 55809 2148.9
- bmi      1    1852.3 58037 2152.1
- gender   1   10941.8 67126 2216.1
```

Answer:

$$\hat{y}\left(SVR_{bmi=20,gender=female}\right) = 113.67$$

$$\hat{y}\left(SVR_{bmi=20,gender=male}\right) = 125.93$$

$$\hat{y}\left(SVR_{bmi=25,gender=female}\right) = 117.50$$

$$\hat{y}\left(SVR_{bmi=25,gender=male}\right) = 129.76$$

According to the ANOVA comparing models 1 (gender + bmi) and 2 (gender * bmi), the model using gender and BMI without interaction (model 1) better predicts SVR. Instead of doing separate models for different combinations of gender, BMI, sport and smoking I decided to perform a stepwise regression model to verify whether sport or smoking are relevant in predicting SVR. I concluded that the better model to predict SVR continues to be gender + bmi.