2nd Assignment: Probability and Statistics for Data Analysis

Code ▾

1. In file “data.txt” (available on the e-class assignments site), you willfind the recorded variables Y, X1, X2, X3, X4 (continuous), and W (categorical with three levels) on 150 cases. Using these data, answer the following questions

Hide

```
data <- read.delim("data.txt", header = TRUE, quote = "\"",, sep = " ")
View(data)

weightloss <- read.delim("weightloss.txt", header = TRUE, quote = "\"",, sep = " ")
View(weightloss)
```

(a) Run the parametric one-way ANOVA of each of the continuous variables (Y, X1, X2, X3, X4) on the categorical variable (W). Specifically,

(i) provides a graphical representation of each of the continuous versus the categorical variable

Hide

```
unique_W <- unique(data$W)
cat("Unique W Categories:", unique_W, "\n")
```

```
Unique W Categories: A C B
```

Hide

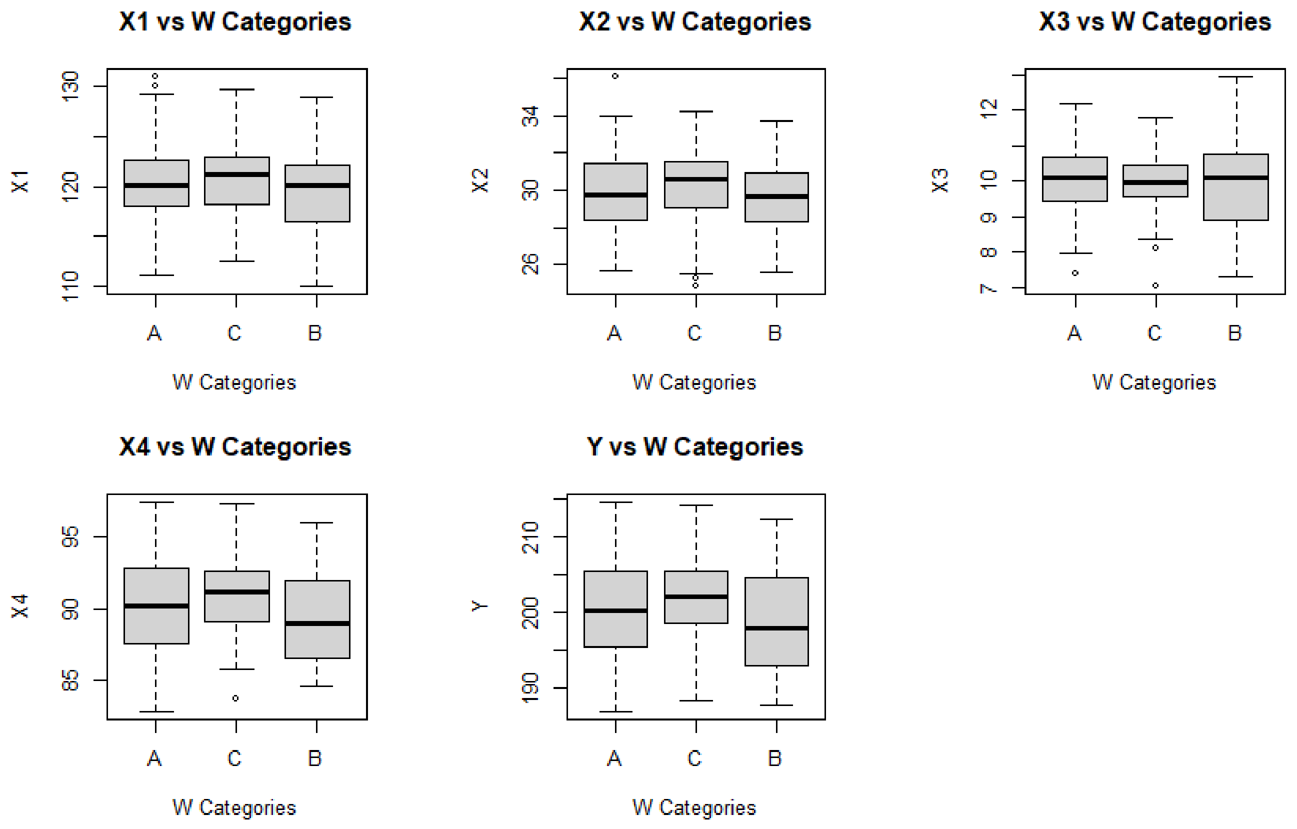
```
par(mfrow = c(2, 3))
par(mar = c(4, 4 ,4, 4))
boxplot(X1~factor(W),data=data,
        main=paste("X1", "vs W Categories"),
        xlab="W Categories", ylab="X1",
        names=c("A","C","B"))

boxplot(X2~factor(W),data=data,
        main=paste("X2", "vs W Categories"),
        xlab="W Categories", ylab="X2",
        names=c("A","C","B"))
```

Hide

```
boxplot(X3~factor(W),data=data,
        main=paste("X3", "vs W Categories"),
        xlab="W Categories", ylab="X3",
        names=c("A","C","B"))
boxplot(X4~factor(W),data=data,
        main=paste("X4", "vs W Categories"),
        xlab="W Categories", ylab="X4",
        names=c("A","C","B"))
```

Hide

```
boxplot(Y~factor(W),data=data,
        main=paste("Y", "vs W Categories"),
        xlab="W Categories", ylab="Y",
        names=c("A","C","B"))
par(mfrow = c(1, 1))
```



(ii) provide the ANOVA output

(iii) check the assumptions.

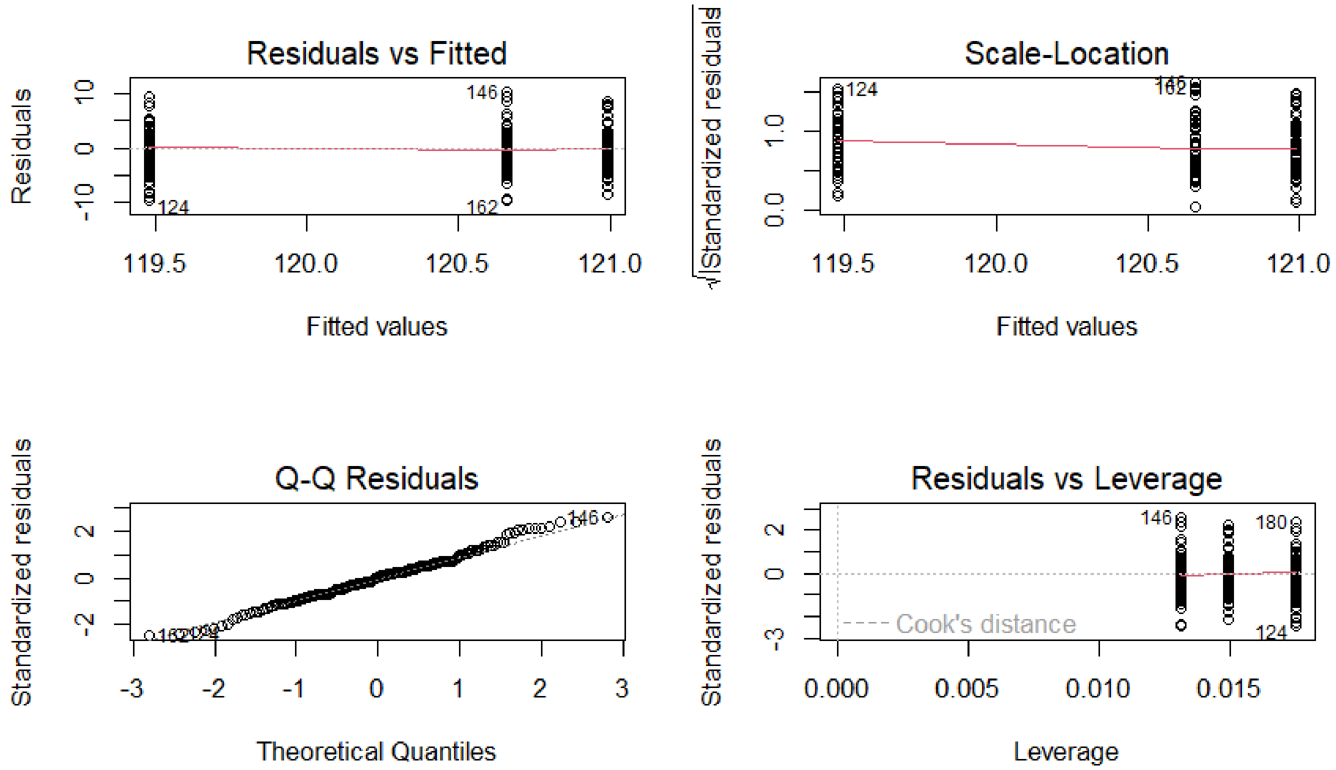
Hide

```
fitx1<-aov(X1~factor(W),data=data)
summary(fitx1)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
factor(W)     2   76.3   38.13    2.42 0.0915 .
Residuals   197 3104.1   15.76
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
layout(matrix(1:4,2,2))
plot(fitx1)
```



Hide

```
shapiro.test(fitx1$residuals)
```

```

	Shapiro-Wilk normality test

data:  fitx1$residuals
W = 0.99123, p-value = 0.268
```

Hide

```
bartlett.test(X1~factor(W),data=data)
```

```

	Bartlett test of homogeneity of variances

data:  X1 by factor(W)
Bartlett's K-squared = 1.8261, df = 2, p-value = 0.4013
```

Hide

```
fligner.test(X1~factor(W),data=data)
```

```

	Fligner-Killeen test of homogeneity of variances

data:  X1 by factor(W)
Fligner-Killeen:med chi-squared = 2.3293, df = 2, p-value = 0.312
```

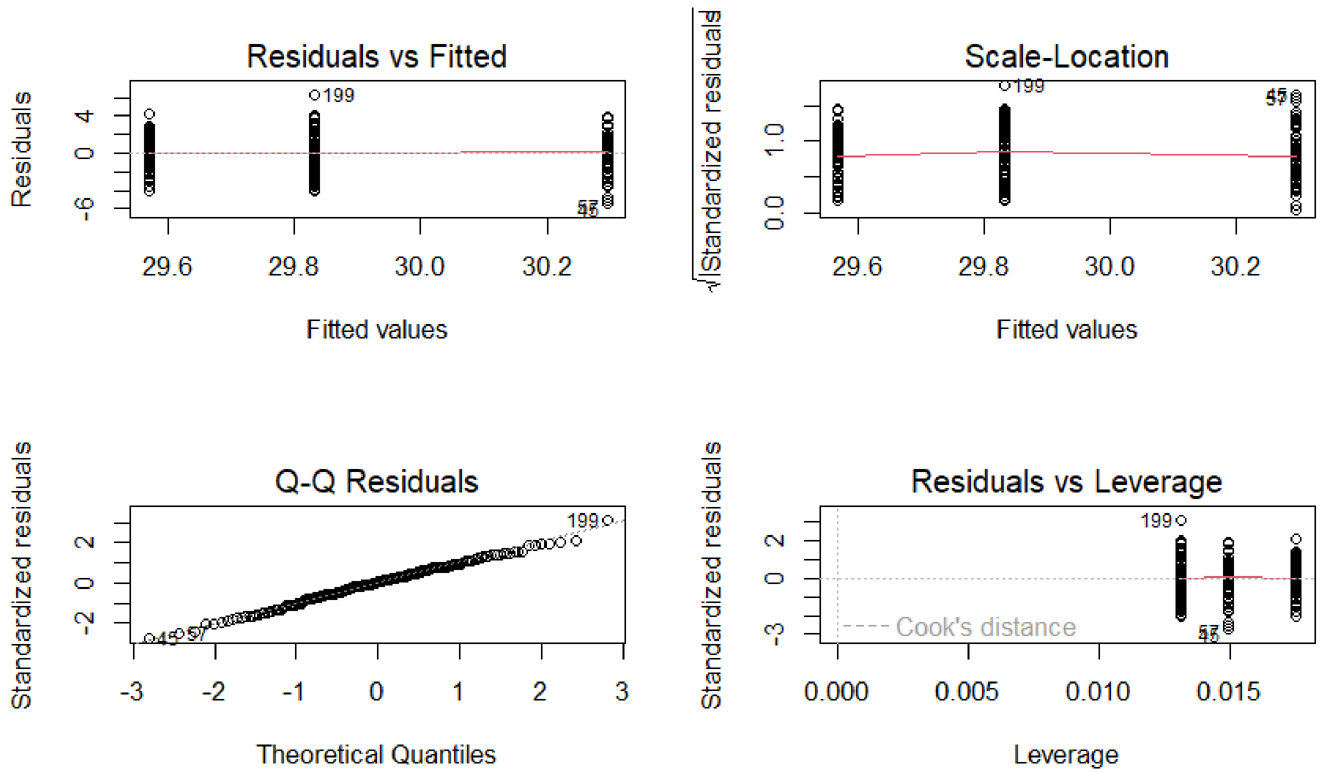
Hide

```
fitx2<-aov(X2~factor(W),data=data)
summary(fitx2)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
factor(W)     2   17.0   8.489   2.079  0.128
Residuals   197  804.3   4.083
```

Hide

```
layout(matrix(1:4,2,2))
plot(fitx2)
```



Hide

```
shapiro.test(fitx2$residuals)
```

```
	Shapiro-Wilk normality test

data:  fitx2$residuals
W = 0.99539, p-value = 0.8049
```

Hide

```
bartlett.test(X2~factor(W),data=data)
```

```
	Bartlett test of homogeneity of variances

data:  X2 by factor(W)
Bartlett's K-squared = 2.2362, df = 2, p-value = 0.3269
```

Hide

```
fligner.test(X2~factor(W),data=data)
```

```
	Fligner-Killeen test of homogeneity of variances

data:  X2 by factor(W)
Fligner-Killeen:med chi-squared = 2.1984, df = 2, p-value = 0.3331
```

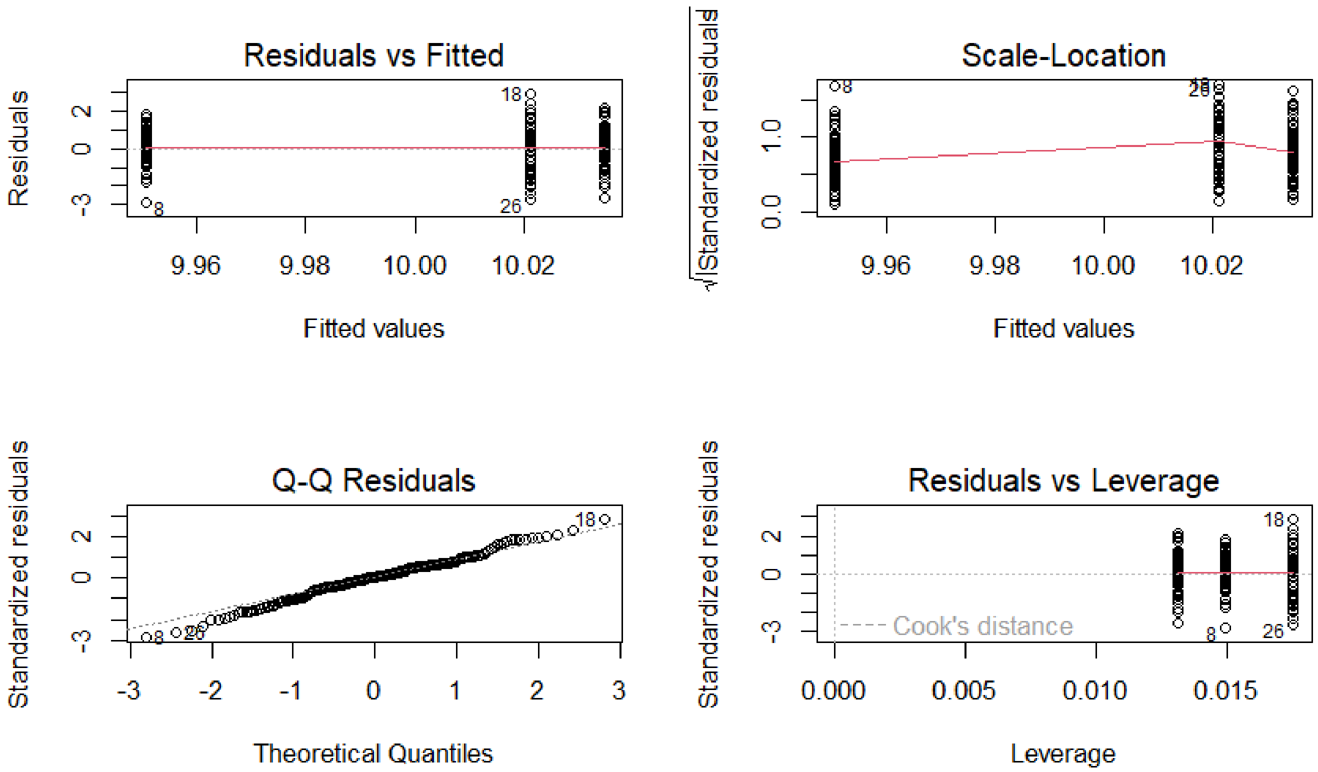
Hide

```
fitx3<-aov(X3~factor(W),data=data)
summary(fitx3)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
factor(W)     2   0.28  0.1397   0.133  0.876
Residuals   197 207.24  1.0520
```

Hide

```
layout(matrix(1:4,2,2))
plot(fitx3)
```



Hide

```
shapiro.test(fitx3$residuals)
```

```

	Shapiro-Wilk normality test

data:  fitx3$residuals
W = 0.99108, p-value = 0.2555
```

Hide

```
bartlett.test(X3~factor(W),data=data)
```

```
	Bartlett test of homogeneity of variances

data:  X3 by factor(W)
Bartlett's K-squared = 14.682, df = 2, p-value = 0.0006485
```

Hide

```
fligner.test(X3~factor(W),data=data)
```

```
	Fligner-Killeen test of homogeneity of variances

data:  X3 by factor(W)
Fligner-Killeen:med chi-squared = 12.686, df = 2, p-value = 0.001759
```

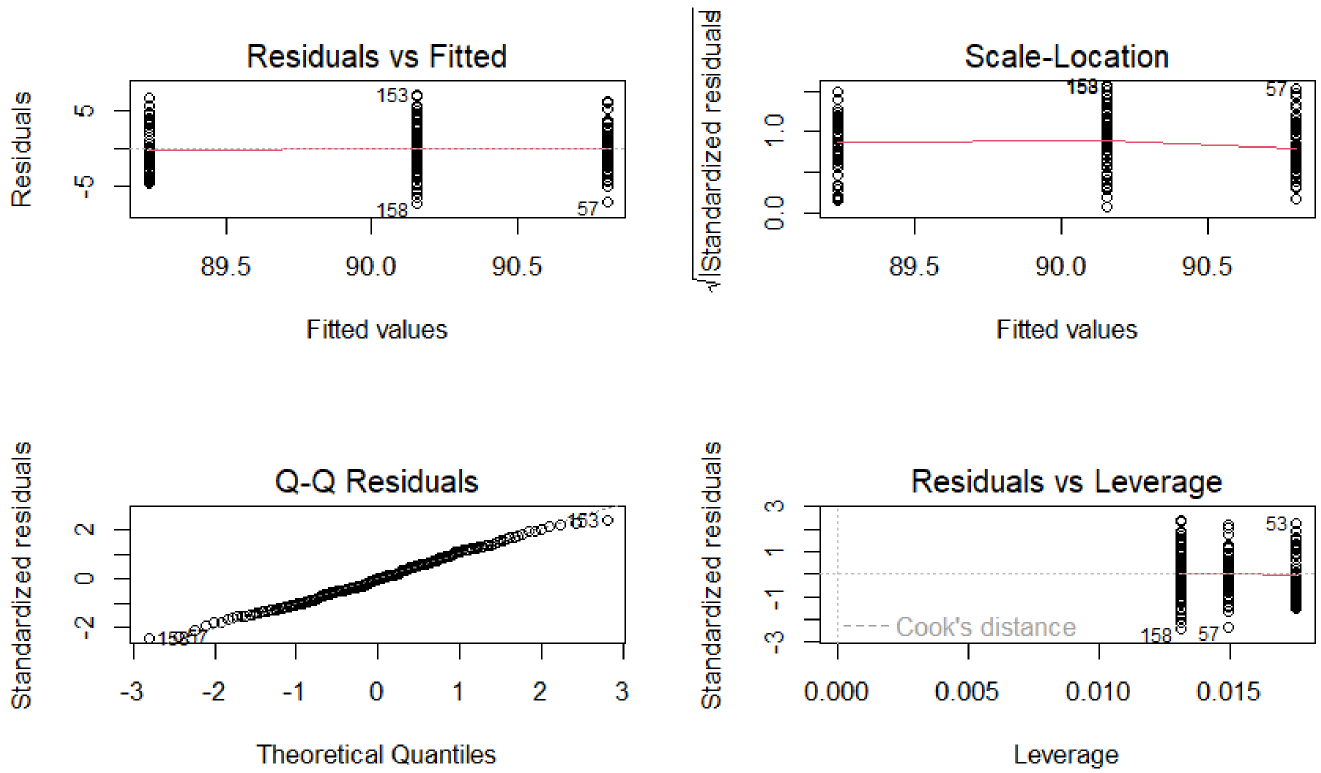
Hide

```
fitx4<-aov(X4~factor(W),data=data)
summary(fitx4)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)  
factor(W)     2   75.8   37.89   4.171 0.0168 *
Residuals   197 1789.6    9.08                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
layout(matrix(1:4,2,2))
plot(fitx4)
```



Hide

```
shapiro.test(fitx4$residuals)
```

```

    Shapiro-Wilk normality test

data:  fitx4$residuals
W = 0.99272, p-value = 0.4243
```

Hide

```
bartlett.test(X4~factor(W),data=data)
```

```

    Bartlett test of homogeneity of variances

data:  X4 by factor(W)
Bartlett's K-squared = 4.0286, df = 2, p-value = 0.1334
```

Hide

```
fligner.test(X4~factor(W),data=data)
```

```

    Fligner-Killeen test of homogeneity of variances

data:  X4 by factor(W)
Fligner-Killeen:med chi-squared = 4.3185, df = 2, p-value = 0.1154
```

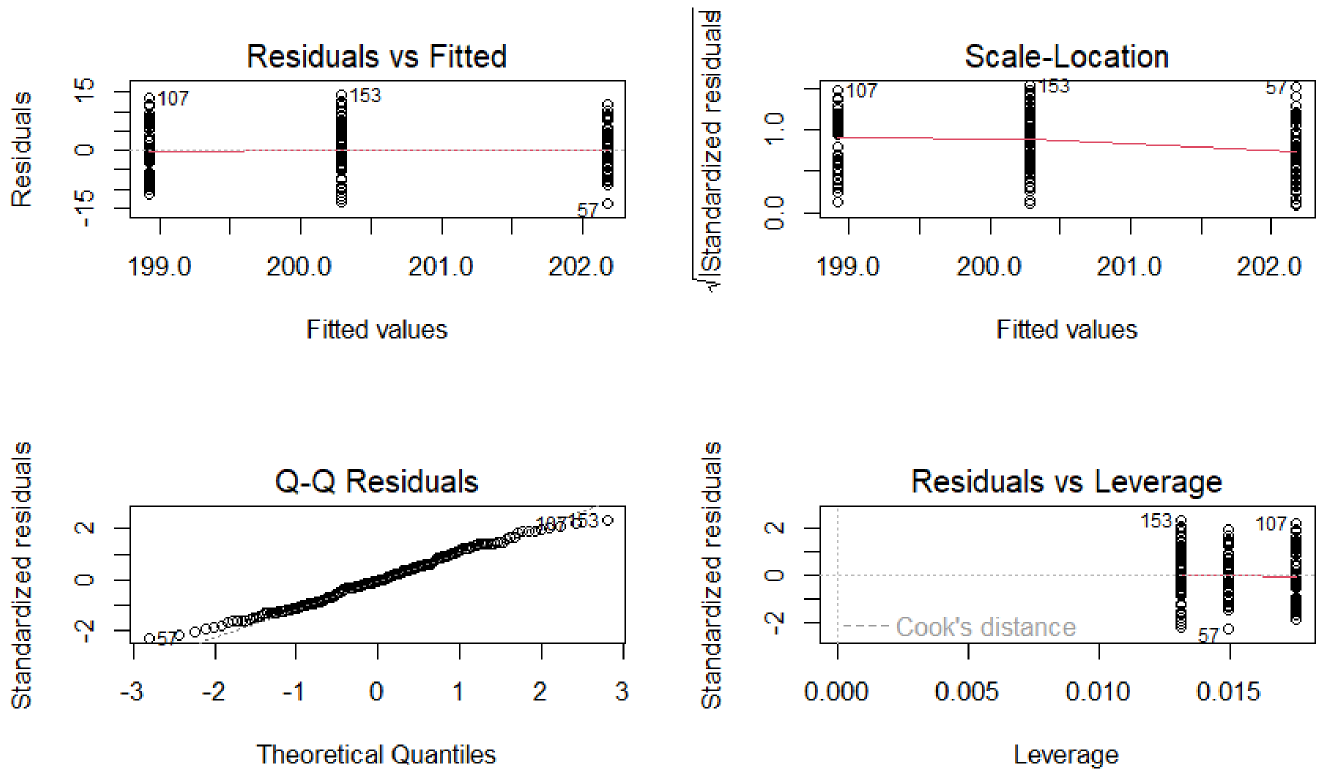
Hide

```
fity<-aov(Y~factor(W),data=data)
summary(fity)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)  
factor(W)     2    333  166.71   4.352 0.0141 *
Residuals   197   7546   38.31                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
layout(matrix(1:4,2,2))
plot(fity)
```



Hide

```
shapiro.test(fity$residuals)
```

```

    Shapiro-Wilk normality test

data:  fity$residuals
W = 0.98923, p-value = 0.1374
```

Hide

```
bartlett.test(Y~factor(W),data=data)
```

```
	Bartlett test of homogeneity of variances

data:  Y by factor(W)
Bartlett's K-squared = 4.6956, df = 2, p-value = 0.09558
```

Hide

```
fligner.test(Y~factor(W),data=data)
```

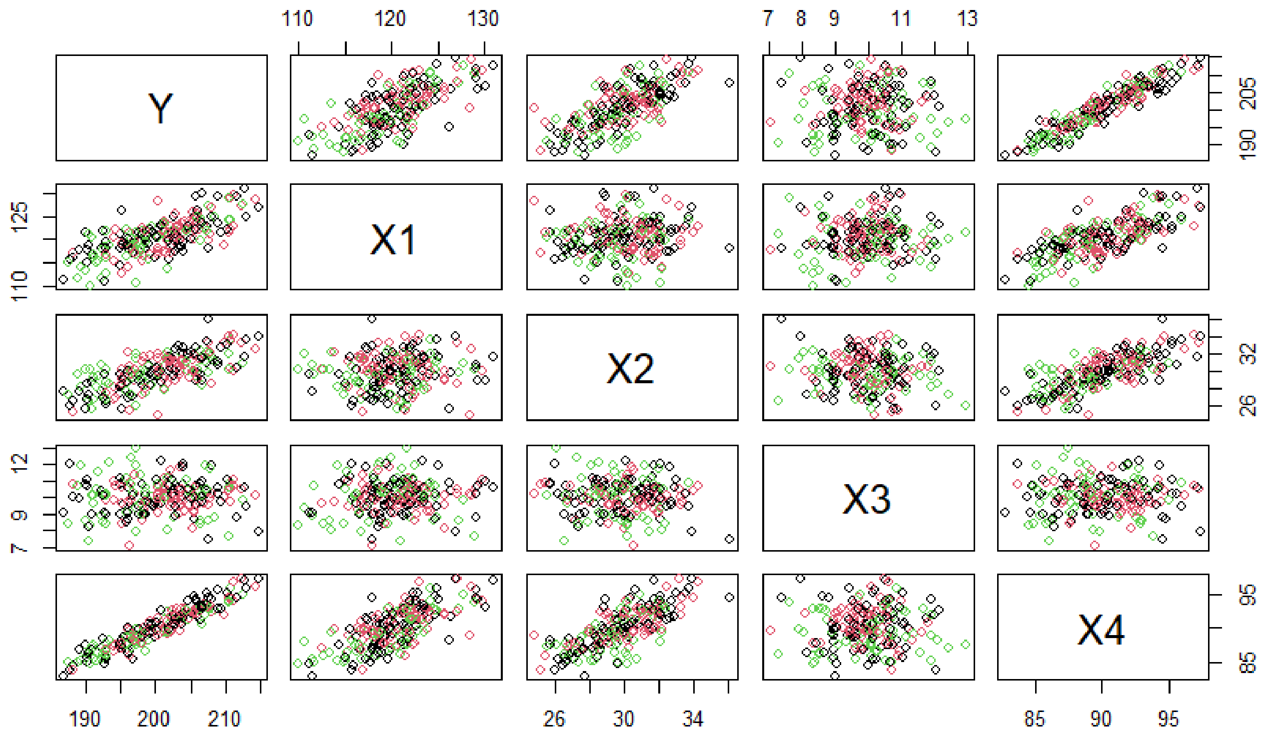
```
	Fligner-Killeen test of homogeneity of variances

data:  Y by factor(W)
Fligner-Killeen:med chi-squared = 7.0448, df = 2, p-value = 0.02953
```

As we can see above, foe all 5 models the residuals are normally distributed, with constant variance and mean value close to 0. So the assumptions exist in our models.

(b) Provide a scatter-plot matrix of Y, X1, X2, X3, and X4, annotatingthe different levels of W in each plot using a different color.

Hide

```
data$W <- as.factor(data$W)
pairs(data[, c("Y", "X1", "X2", "X3", "X4")], col = as.numeric(data$W))
legend("topright", legend = levels(data$W), col = 1:length(levels(data$W)), pch = 1, title =
"W")
```



(c) Run the regression model of Y on X4

Hide

```
model <- lm(Y ~ X4, data = data)
summary(model)
```

```
Call:
lm(formula = Y ~ X4, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.5133 -1.3818  0.1039  1.4803  5.9044 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  26.1973     4.4449   5.894  1.6e-08 ***
X4            1.9347     0.0493  39.243  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.129 on 198 degrees of freedom
Multiple R-squared:  0.8861,	Adjusted R-squared:  0.8855 
F-statistic:  1540 on 1 and 198 DF,  p-value: < 2.2e-16
```

(d) Run the regression model of Y on all the remaining variables (X1,X2, X3, X4, W), including the non-additive terms (i.e., interactions of the continuous predictors with the categorical).

Hide

```
model2 <- lm(Y ~ X1 + X2 + X3 + X4 + W + X1:W + X2:W + X3:W + X4:W, data = data)
summary(model2)
```

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + W + X1:W + X2:W + X3:W + 
    X4:W, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8807 -1.3656 -0.0337  1.0723  5.4653 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  28.3612     7.1589   3.962 0.000106 ***
X1            1.1682     0.2570   4.545 9.90e-06 ***
X2            2.7008     0.5276   5.119 7.64e-07 ***
X3            0.3221     0.2313   1.393 0.165391    
X4           -0.5859     0.5015  -1.168 0.244184    
WB           -8.2392    11.6561  -0.707 0.480544    
WC          -24.4132    10.7774  -2.265 0.024658 *  
X1:WB        -0.2119     0.3432  -0.617 0.537741    
X1:WC        -0.4392     0.3618  -1.214 0.226304    
X2:WB        -0.9233     0.7186  -1.285 0.200463    
X2:WC        -1.3562     0.7368  -1.841 0.067257 .  
X3:WB         0.2838     0.3743   0.758 0.449266    
X3:WC        -0.3090     0.3076  -1.005 0.316355    
X4:WB         0.6572     0.6797   0.967 0.334848    
X4:WC         1.3478     0.7030   1.917 0.056730 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.879 on 185 degrees of freedom
Multiple R-squared:  0.9171,    Adjusted R-squared:  0.9108 
F-statistic: 146.2 on 14 and 185 DF,  p-value: < 2.2e-16
```

(e) Examine the regression assumptions and provide alternatives if any of them fails.

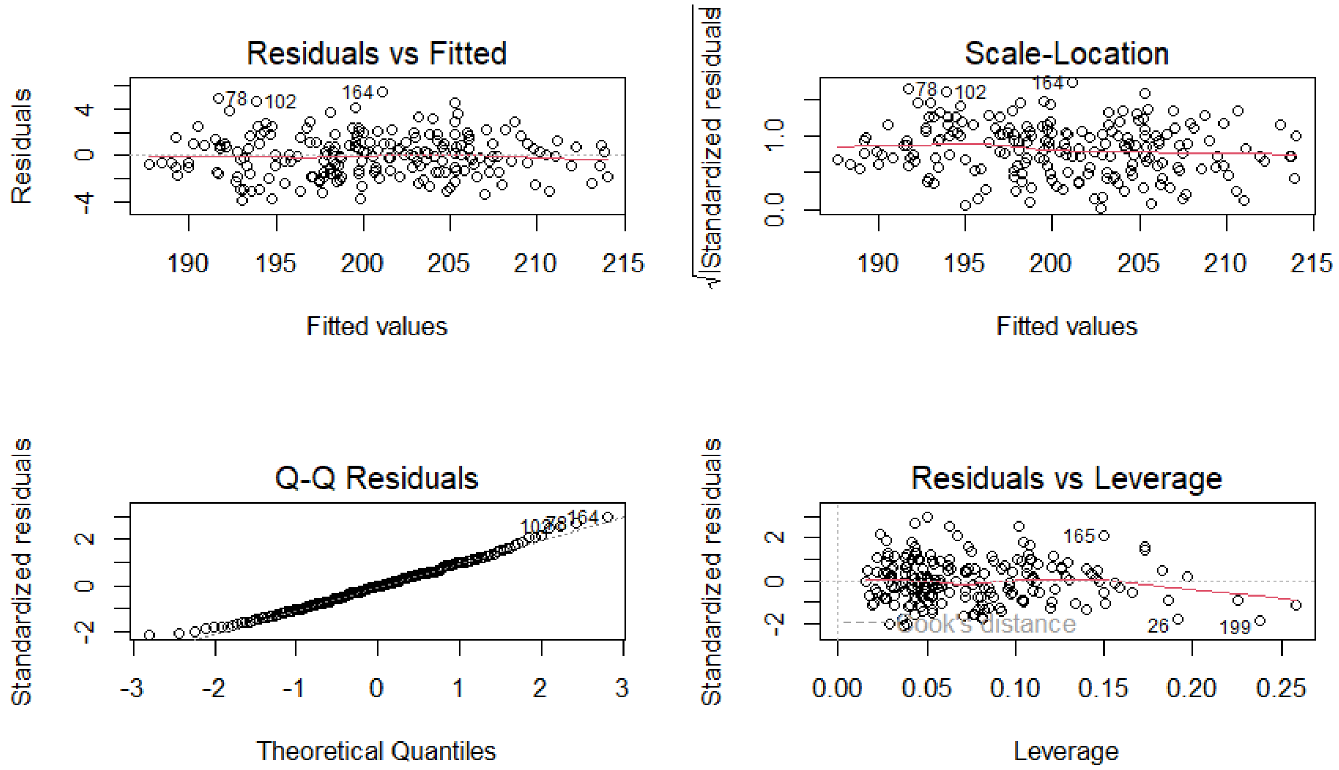
Hide

```
shapiro.test(model2$residuals)
```

```
	Shapiro-Wilk normality test

data:  model2$residuals
W = 0.9907, p-value = 0.2253
```

Hide

```
layout(matrix(1:4,2,2))
plot(model2)
```



The residuals of our model seem to be normal, with constant variance and mean equal to 0, so the assumptions of our model exist.

(f) Use the “stepwise regression” approach to examine whether you can reduce the dimension of the model.

Hide

```
fitnull<-lm(Y ~ 1,data=data)
stepSR<-step(fitnull, scope=list(lower = ~ 1,
                                  upper = ~ X1 + X2 + X3 + X4 + W + X1:W + X2:W + X3:W + X4:
W),
          direction="both", data=data)
```

```
Start:  AIC=736.75
Y ~ 1

       Df Sum of Sq    RSS    AIC
+ X4    1    6982.1  897.7 304.30
+ X2    1    3966.9 3912.8 598.74
+ X1    1    3892.7 3987.0 602.50
+ W     2     333.4 7546.3 732.10
<none>              7879.7 736.75
+ X3    1       6.2 7873.5 738.59

Step:  AIC=304.3
Y ~ X4

       Df Sum of Sq    RSS    AIC
+ X1    1      11.6  886.1 303.70
<none>               897.7 304.30
+ W     2      14.8  882.9 304.97
+ X2    1       1.0  896.7 306.09
+ X3    1       0.8  896.9 306.13
- X4    1    6982.1 7879.7 736.75

Step:  AIC=303.7
Y ~ X4 + X1

       Df Sum of Sq    RSS    AIC
+ X2    1    145.70  740.4 269.78
<none>               886.1 303.70
+ W     2     16.10  870.0 304.04
- X1    1     11.58  897.7 304.30
+ X3    1      0.31  885.8 305.63
- X4    1   3100.94 3987.0 602.50

Step:  AIC=269.78
Y ~ X4 + X1 + X2

       Df Sum of Sq    RSS    AIC
- X4    1     1.293 741.70 268.12
+ X3    1    10.964 729.44 268.79
<none>              740.40 269.78
+ W     2    14.393 726.01 269.85
- X2    1   145.696 886.10 303.70
- X1    1   156.321 896.73 306.09

Step:  AIC=268.12
Y ~ X1 + X2

       Df Sum of Sq    RSS    AIC
+ X3    1      12.1  729.6 266.84
<none>               741.7 268.12
+ W     2      14.4  727.3 268.19
+ X4    1       1.3  740.4 269.78
- X1    1    3171.1 3912.8 598.74
- X2    1    3245.3 3987.0 602.50
```

```
Step:  AIC=266.84
Y ~ X1 + X2 + X3

       Df Sum of Sq    RSS    AIC
+ W     2      15.2  714.5 266.64
<none>               729.6 266.84
- X3    1      12.1  741.7 268.12
+ X4    1       0.2  729.4 268.79
- X1    1    3111.9 3841.6 597.06
- X2    1    3251.7 3981.3 604.21

Step:  AIC=266.64
Y ~ X1 + X2 + X3 + W

       Df Sum of Sq    RSS    AIC
+ X1:W  2     27.66  686.8 262.74
<none>               714.5 266.64
- W     2     15.18  729.6 266.84
+ X2:W  2     11.42  703.0 267.42
- X3    1     12.81  727.3 268.19
+ X3:W  2      8.57  705.9 268.23
+ X4    1      0.15  714.3 268.60
- X1    1   3037.25 3751.7 596.33
- X2    1   3160.70 3875.2 602.80

Step:  AIC=262.74
Y ~ X1 + X2 + X3 + W + X1:W

       Df Sum of Sq    RSS    AIC
<none>               686.8 262.74
+ X2:W  2     12.73  674.1 263.00
+ X3:W  2      9.57  677.2 263.94
- X3    1     12.53  699.3 264.36
+ X4    1      0.14  686.7 264.70
- X1:W  2     27.66  714.5 266.64
- X2    1   3149.67 3836.5 604.80
```

The dimensionallity of the model was indeed reduced, from 9 independent variables to 5

(g) Using the model found in (f), provide a point estimate and a 95% confidence interval for the prediction of Y when: (X1, X2, X3, X4, W) = (120, 30, 10, 90, B)

Hide

```
step_model = lm(Y ~ X1 +X2 + X3 + W + X1:W,data=data)
summary(step_model)
```

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + W + X1:W, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.2069 -1.2826  0.0146  1.1169  5.6274 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  30.78245    6.87991   4.474 1.31e-05 ***
X1            0.88574    0.05643  15.696  < 2e-16 ***
X2            2.01609    0.06794  29.674  < 2e-16 ***
X3            0.24803    0.13252   1.872  0.06278 .  
WB          -13.51907   10.31593  -1.311  0.19159    
WC          -26.58912    9.69468  -2.743  0.00667 ** 
X1:WB         0.11739    0.08529   1.376  0.17032    
X1:WC         0.22430    0.08072   2.779  0.00600 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.891 on 192 degrees of freedom
Multiple R-squared:  0.9128,	Adjusted R-squared:  0.9097 
F-statistic: 287.3 on 7 and 192 DF,  p-value: < 2.2e-16
```

Hide

```
new_data <- data.frame(X1 = 120, X2 = 30, X3 = 10, W = "B")
prediction_with_interval <- predict(step_model, newdata = new_data, interval = "confidence")
predicted_Y <- prediction_with_interval[1]
lower_bound <- prediction_with_interval[2]
upper_bound <- prediction_with_interval[3]
cat("Predicted Y:", predicted_Y, "\n")
```

```
Predicted Y: 200.6028
```

Hide

```
cat("95% Confidence Interval:", lower_bound, "to", upper_bound, "\n")
```

```
95% Confidence Interval: 200.1282 to 201.0774
```

(h) Using the cut() function, create a categorical variable (named Z) with three levels based on the quantiles of X4. Provide the contingency table of X4 and W.

Hide

```
data$Z <- cut(data$X4, breaks = quantile(data$X4, probs = c(0, 1/3, 2/3, 1)), labels = c("Lev
el1", "Level2", "Level3"))
contingency_table <- table(data$X4, data$W)
print(contingency_table)
```

```
      A B C
82.86 1 0 0
83.74 0 1 0
83.8  1 0 0
84.42 1 0 0
84.63 0 0 1
84.74 1 0 0
84.83 1 0 0
84.95 0 0 1
84.98 0 0 1
85.16 0 0 1
85.18 0 0 1
85.36 0 0 1
85.39 1 0 0
85.45 0 0 1
85.47 1 0 0
85.61 1 0 0
85.68 1 0 0
85.73 0 0 1
85.81 0 0 1
85.84 0 1 0
86.06 0 0 1
86.13 0 0 1
86.31 0 0 1
86.33 1 0 0
86.36 0 0 1
86.4  1 0 0
86.44 0 0 1
86.45 0 1 0
86.51 0 0 1
86.58 1 0 0
86.73 0 1 0
86.75 1 0 0
86.79 1 0 0
86.93 1 0 0
87.01 0 1 0
87.09 1 0 0
87.12 0 0 1
87.2  0 1 0
87.23 0 0 1
87.43 1 0 0
87.44 1 0 0
87.46 0 0 1
87.5  0 1 0
87.52 1 0 0
87.59 0 0 1
87.6  1 0 1
87.63 0 1 0
87.68 1 0 0
87.71 0 1 0
87.74 0 0 1
87.8  0 0 1
87.81 0 1 0
87.87 1 0 0
```

```
87.97 0 0 1
88.19 0 0 1
88.23 1 0 0
88.26 0 1 0
88.3  0 0 1
88.35 1 0 0
88.58 0 0 1
88.62 0 0 1
88.63 1 0 0
88.65 0 1 0
88.66 1 0 0
88.67 1 0 0
88.76 0 1 0
88.84 0 1 0
88.9  0 0 1
88.91 1 0 0
88.92 0 0 1
88.93 0 1 0
88.96 0 1 0
88.99 0 1 0
89    1 0 0
89.04 1 0 0
89.14 1 0 0
89.16 0 0 1
89.18 0 1 0
89.24 1 0 0
89.25 0 1 0
89.31 0 0 1
89.32 0 0 1
89.35 0 0 1
89.36 0 1 0
89.37 0 0 1
89.45 1 0 0
89.49 0 0 1
89.5  0 2 0
89.57 0 1 0
89.58 0 1 0
89.59 0 0 1
89.66 0 1 0
89.74 1 1 0
89.76 0 1 0
89.86 1 0 0
89.88 0 0 1
89.89 0 1 0
89.9  1 0 0
89.91 1 0 0
89.95 0 1 0
90.06 0 1 0
90.14 1 0 0
90.16 0 1 0
90.27 1 0 0
90.28 0 1 0
90.43 0 0 1
90.45 1 0 0
90.48 1 0 0
90.53 1 0 0
```

```
90.55 1 0 0
90.57 1 0 0
90.64 1 0 0
90.72 0 1 0
90.77 1 0 0
91.1  0 1 0
91.12 0 1 0
91.15 1 0 0
91.18 1 0 0
91.22 0 0 2
91.23 0 0 1
91.26 0 1 0
91.3  1 0 0
91.33 0 1 0
91.36 0 1 0
91.47 0 1 0
91.57 1 0 0
91.69 1 0 0
91.71 0 1 1
91.78 0 1 0
91.81 1 0 0
91.85 0 2 0
91.91 2 0 0
91.93 0 0 1
91.96 0 0 1
92.17 0 1 0
92.2  1 0 0
92.21 0 1 0
92.25 0 2 0
92.27 1 0 0
92.29 0 1 0
92.36 0 1 0
92.47 0 0 1
92.48 0 1 0
92.57 0 0 1
92.58 0 1 0
92.61 0 1 0
92.62 0 1 0
92.66 0 0 1
92.72 1 0 0
92.75 0 0 1
92.78 1 0 0
92.79 0 0 1
92.87 1 0 0
92.9  0 0 1
92.92 0 1 0
92.93 1 0 1
92.95 0 1 0
92.96 0 1 0
92.99 0 0 1
93.01 0 1 0
93.16 1 0 0
93.34 0 0 1
93.39 0 1 0
93.44 0 1 0
93.46 1 0 0
```

```
  93.52 1 0 0
  93.65 0 1 0
  93.85 0 0 1
  94.02 1 0 0
  94.13 0 0 1
  94.15 1 0 0
  94.16 0 1 0
  94.23 0 1 0
  94.29 1 0 0
  94.33 1 0 0
  94.42 0 1 0
  94.54 1 0 0
  94.56 1 0 0
  94.62 0 1 0
  94.78 1 0 0
  94.98 0 0 1
  95.17 1 0 0
  95.18 1 0 0
  95.6  1 0 0
  95.93 1 0 1
  96.11 0 1 0
  96.91 0 1 0
  97.06 1 0 0
  97.26 0 1 0
  97.36 1 0 0
```

(i) Run the parametric two-way ANOVA of Y on the categorical variables W and Z (including the interaction term). Provide the fit, examine the assumptions, and comment on the significance of the terms.

Hide

```
model2wayAnova <- aov(Y ~ W * Z, data = data)
summary(model2wayAnova)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
W             2    328   164.2  19.102 2.76e-08 ***
Z             2   5704  2852.0 331.883  < 2e-16 ***
W:Z           4     28     6.9   0.808    0.521    
Residuals   190   1633     8.6                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
1 observation deleted due to missingness
```

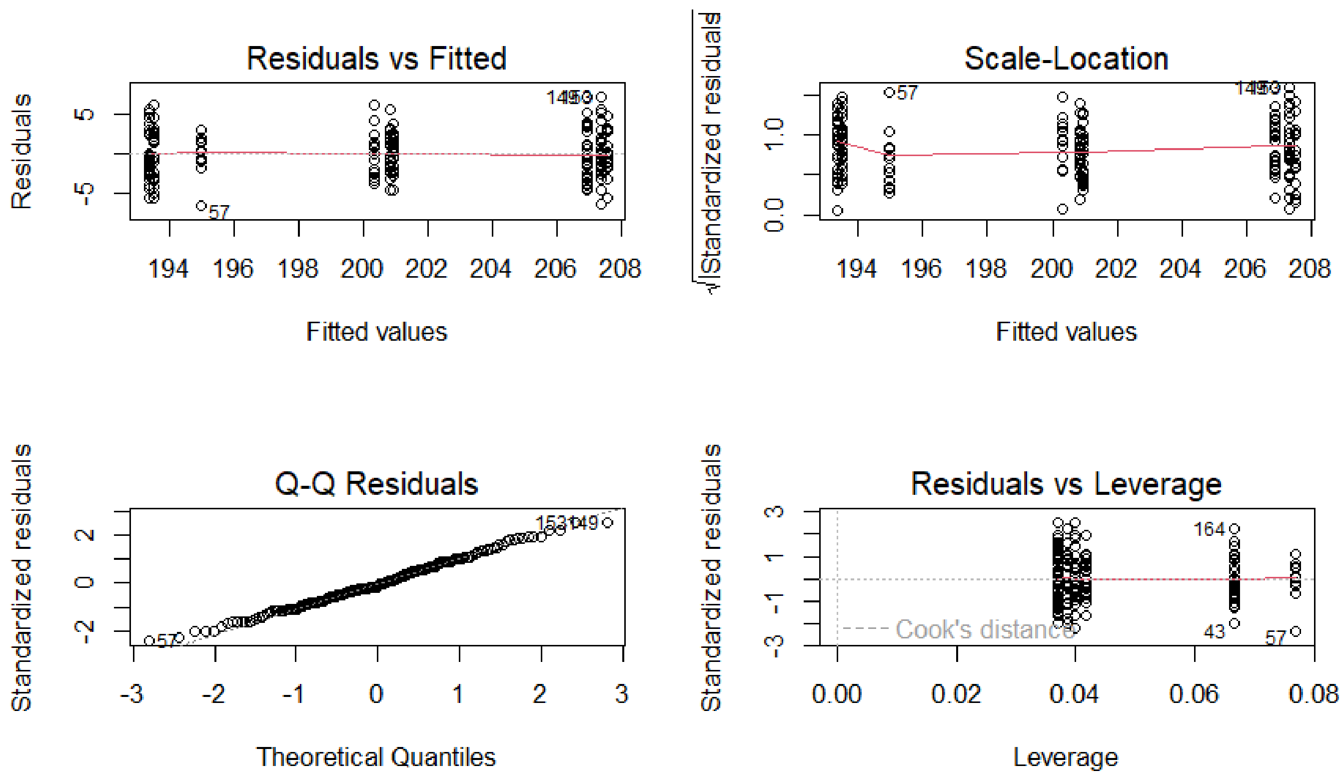
Hide

```
shapiro.test(model2wayAnova$residuals)
```

```
    Shapiro-Wilk normality test

data:  model2wayAnova$residuals
W = 0.99288, p-value = 0.4493
```

Hide

```
layout(matrix(1:4,2,2))
plot(model2wayAnova)
```



The residuals of our model are normal with constant variance and 0 mean. Both W and Z variables were statistically significant for our ANOVA model, but the interaction between them wasn't. That could mean that our dependent variable Y has a greater variance due to W and Z variables, but we can't be certain. For a different model the significance values could be different. Still, the p-values for these categorical variables were almost 0, meaning that they were very significant for our model.

2. In the file "weightloss.txt"(available on the e-class assignments site)you will find the recorded variables work (categorical with three levels),diet (categorical with four levels), and loss (continue, in calories). More specifically, the data provide the weight loss per day in a 3×4 factorial experiment. The

two factors include 3 types of workout and 4 types of diet. Each combination of the two factors is used to be completely randomized.

(a) Provide boxplots of the weight loss per workout, per diet, and for the combinations of the two categorical factors.

Hide

```
unique_work <- unique(weightloss$workout)
cat("Unique Workout Categories:", unique_work, "\n")
```

```
Unique Workout Categories: W3 W1 W2
```

Hide

```
unique_d <- unique(weightloss$diet)
cat("Unique Diet Categories:", unique_d, "\n")
```

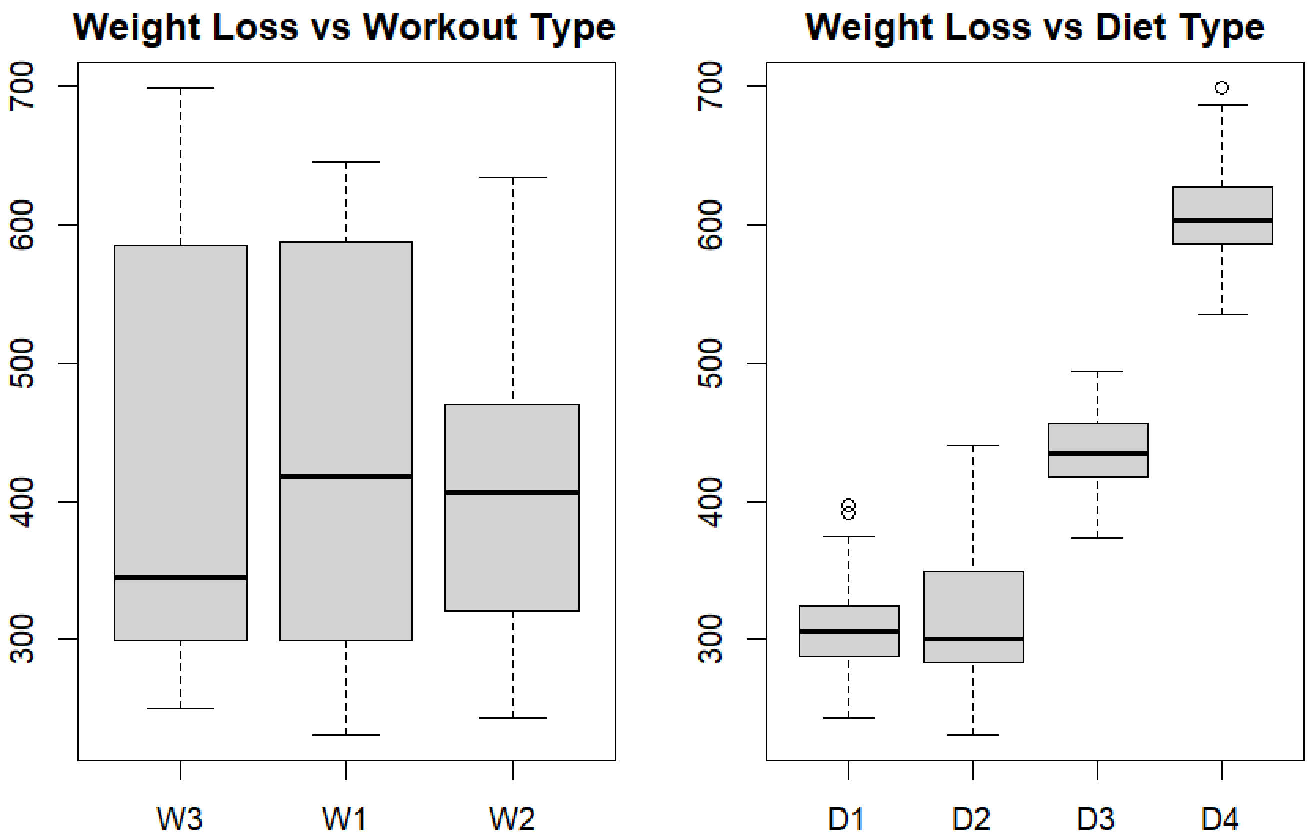
```
Unique Diet Categories: D1 D2 D3 D4
```

Hide

```
par(mfrow = c(1, 2))
par(mar = c(2, 2 , 2, 2))
boxplot(loss~factor(workout),data=weightloss,
        main=paste("Weight Loss", "vs Workout Type"),
        xlab="Workout Types", ylab="Loss",
        names=c("W3","W1","W2"))
boxplot(loss~factor(diet),data=weightloss,
        main=paste("Weight Loss", "vs Diet Type"),
        xlab="Workout Types", ylab="Loss",
        names=c("D1","D2","D3","D4"))
```

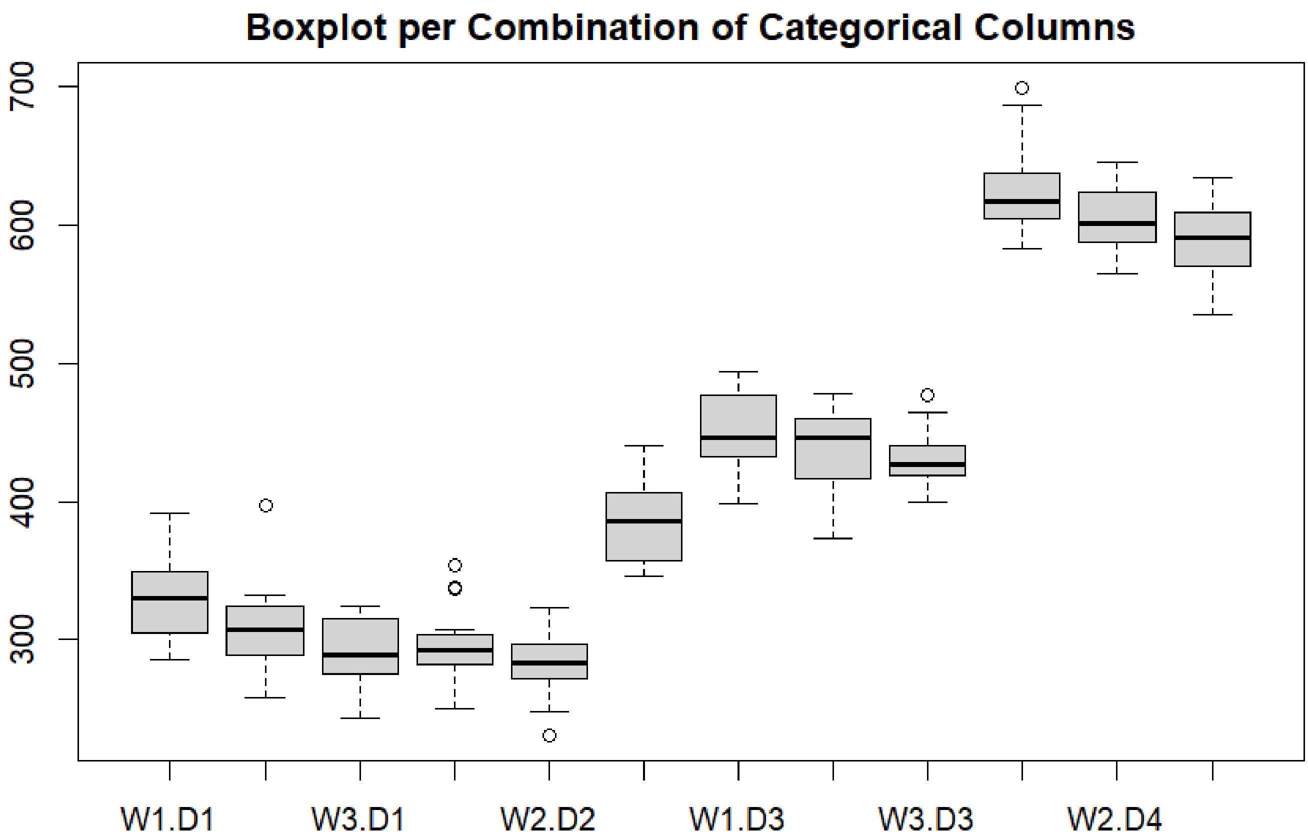
Hide

```


weightloss$combinations <- interaction(weightloss$workout, weightloss$diet)
unique_combinations <- levels(weightloss$combinations)
unique_combinations_array <- as.array(unique_combinations)

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



Hide

```
boxplot(loss ~ combinations, data=weightloss,
        main = "Boxplot per Combination of Categorical Columns",
        xlab = "Combination",
        ylab = "Loss",
        names=unique_combinations_array)
```



(b) Fit a One-Way ANOVA model with the weight loss as a response and the workout (as a factor). Interpret the model parameters.

Hide

```
model_work <- aov(loss ~ workout, data = weightloss)
summary(model_work)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)
workout       2   13831    6916     0.4  0.671
Residuals   237 4101316   17305
```

The most important parameters we have to interpret in this model are the mean sum of squares for the workout variable and the residuals, to see which of these have a great effect. As we can see the mean SoS for our variable workout is very low comparing the residuals one (very high p-value of the F-Statistic) meaning that our dependent variable Y doesn’t vary depending to the workout but more on other effects.

(c) In the ANOVA model of (b), is the expected difference between W2 and W3 significant? [TIP: change the reference level appropriately and refit the ANOVA model of question (b)]

Hide

```
weightloss$workout <- factor(weightloss$workout)
weightloss$workout <- relevel(weightloss$workout, ref = "W3")
model_ref_W3 <- aov(loss ~ workout, data = weightloss)
summary(model_ref_W3)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)
workout       2   13831    6916     0.4  0.671
Residuals   237 4101316   17305
```

Hide

```
weightloss$workout <- factor(weightloss$workout)
weightloss$workout <- relevel(weightloss$workout, ref = "W2")
model_ref_W2 <- aov(loss ~ workout, data = weightloss)
summary(model_ref_W2)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)
workout       2   13831    6916     0.4  0.671
Residuals   237 4101316   17305
```

We can see that both models with reference to W2 or W3 are the same.

(d) Fit a One-Way ANOVA model for the weight loss as response and diet. Interpret the model parameters. Are all treatments significant?

Hide

```
model_diet <- aov(loss ~ diet, data = weightloss)
summary(model_diet)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)    
diet          3 3803266 1267755   959.3 <2e-16 ***
Residuals   236  311882    1322                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
TukeyHSD(model_diet)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = loss ~ diet, data = weightloss)

$diet
            diff        lwr       upr     p adj
D2-D1   9.411629  -7.419278  26.24254 0.4714878
D3-D1 128.105386 110.036557 146.17421 0.0000000
D4-D1 297.752757 281.371398 314.13412 0.0000000
D3-D2 118.693757 100.437764 136.94975 0.0000000
D4-D2 288.341128 271.753553 304.92870 0.0000000
D4-D3 169.647372 151.804985 187.48976 0.0000000
```

Hide

```
weightloss_excluded <- subset(weightloss, diet != "D1")
model_excluded <- aov(loss ~ diet, data = weightloss_excluded)
summary(model_excluded)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)    
diet          2 2711589 1355795     942 <2e-16 ***
Residuals   173  248982    1439                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that in the Tuckey test, the pair D2-D1 isn’t significant, meaning that for these diets loss varies the same and aren’t both significant for our ANOVA model (same information)

(f) Fit a Two-Way ANOVA model of main effects. Provide the interpretation for the parameters.

Hide

```
model2wayAnova_loss <- aov(loss ~ workout + diet, data = weightloss)
summary(model2wayAnova_loss)
```

```
             Df  Sum Sq Mean Sq F value  Pr(>F)    
workout       2   13831    6916   5.349 0.00535 ** 
diet          3 3798759 1266253 979.329 < 2e-16 ***
Residuals   234  302557    1293                    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hide

```
TukeyHSD(model2wayAnova_loss)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = loss ~ workout + diet, data = weightloss)

$workout
           diff       lwr       upr     p adj
W3-W2 -14.41912 -27.63088 -1.207355 0.0286318
W1-W2 -17.09412 -30.53083 -3.657402 0.0083608
W1-W3  -2.67500 -16.30712 10.957122 0.8886991

$diet
            diff        lwr       upr     p adj
D2-D1   9.108733  -7.540375  25.75784 0.4908204
D3-D1 125.560756 107.687098 143.43441 0.0000000
D4-D1 295.777422 279.573006 311.98184 0.0000000
D3-D2 116.452023  98.393222 134.51082 0.0000000
D4-D2 286.668689 270.260285 303.07709 0.0000000
D4-D3 170.216667 152.567004 187.86633 0.0000000
```

We can see that both diet and workout have significant F-Statistic values, meaning that the mean Sum of Squares for each variable wasn't significantly (or at all) smaller than the one for the Residuals and these categorical values affect the variability of the weight loss.

(g) Exclude the non-significant levels of the factors and refit the model. Provide the interpretation for the parameters of the new, simplified, model.

Hide

```
weightloss_excluded_2 <- subset(weightloss, !(diet == "D2"))
weightloss_excluded_2 <- subset(weightloss_excluded_2, !(workout == "W3"))
model2wayAnova_loss_ex <- aov(loss ~ workout + diet, data = weightloss_excluded_2)
summary(model2wayAnova_loss)
```

```
             Df  Sum Sq Mean Sq F value  Pr(>F)    
workout       2   13831    6916   5.349 0.00535 ** 
diet          3 3798759 1266253 979.329 < 2e-16 ***
Residuals   234  302557    1293                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that the results of the new model are the same as the old, without diet D2, meaning that D1 could by itself explain the variability of the weightloss.

(h) Fit a Two-Way ANOVA model with interactions. Are all the parameters significant

Hide

```
model2wayAnova_loss_IN <- aov(loss ~ workout * diet, data = weightloss)
summary(model2wayAnova_loss_IN)
```

```
              Df  Sum Sq Mean Sq  F value   Pr(>F)    
workout        2   13831    6916    9.836 7.99e-05 ***
diet           3 3798759 1266253 1800.976  < 2e-16 ***
workout:diet   6  142252   23709   33.721  < 2e-16 ***
Residuals    228  160305     703                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
TukeyHSD(model2wayAnova_loss_IN)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = loss ~ workout * diet, data = weightloss)

$workout
           diff       lwr       upr     p adj
W3-W2 -14.41912 -24.16328 -4.674960 0.0016683
W1-W2 -17.09412 -27.00419 -7.184048 0.0001915
W1-W3  -2.67500 -12.72919  7.379188 0.8050786

$diet
            diff        lwr       upr     p adj
D2-D1   9.108733  -3.170853  21.38832 0.2226182
D3-D1 125.560756 112.378000 138.74351 0.0000000
D4-D1 295.777422 283.825819 307.72903 0.0000000
D3-D2 116.452023 103.132714 129.77133 0.0000000
D4-D2 286.668689 274.566634 298.77074 0.0000000
D4-D3 170.216667 157.199119 183.23421 0.0000000

$`workout:diet`
                  diff         lwr        upr     p adj
W3:D1-W2:D1 -16.416667  -43.720202  10.886869 0.7027219
W1:D1-W2:D1  20.924242   -6.906098  48.754583 0.3561075
W2:D2-W2:D1 -24.035088  -52.837173   4.766997 0.2065725
W3:D2-W2:D1  78.611111   49.422404 107.799819 0.0000000
W1:D2-W2:D1 -13.375000  -40.678536  13.928536 0.9010037
W2:D3-W2:D1 129.533333  101.083700 157.982966 0.0000000
W3:D3-W2:D1 122.859649   94.057564 151.661734 0.0000000
W1:D3-W2:D1 143.083333  105.874886 180.291781 0.0000000
W2:D4-W2:D1 297.261905  270.807392 323.716418 0.0000000
W3:D4-W2:D1 280.701754  251.899669 309.503839 0.0000000
W1:D4-W2:D1 317.190476  289.063534 345.317419 0.0000000
W1:D1-W3:D1  37.340909   11.494627  63.187191 0.0002031
W2:D2-W3:D1  -7.618421  -34.508243  19.271401 0.9986765
W3:D2-W3:D1  95.027778   67.724242 122.331314 0.0000000
W1:D2-W3:D1   3.041667  -22.236496  28.319829 0.9999998
W2:D3-W3:D1 145.950000  119.438040 172.461960 0.0000000
W3:D3-W3:D1 139.276316  112.386494 166.166138 0.0000000
W1:D3-W3:D1 159.500000  123.751280 195.248720 0.0000000
W2:D4-W3:D1 313.678571  289.319925 338.037218 0.0000000
W3:D4-W3:D1 297.118421  270.228599 324.008243 0.0000000
W1:D4-W3:D1 333.607143  307.441759 359.772527 0.0000000
W2:D2-W1:D1 -44.959330  -72.383905 -17.534755 0.0000100
W3:D2-W1:D1  57.686869   29.856528  85.517209 0.0000000
W1:D2-W1:D1 -34.299242  -60.145524  -8.452961 0.0010669
W2:D3-W1:D1 108.609091   81.554908 135.663274 0.0000000
W3:D3-W1:D1 101.935407   74.510832 129.359982 0.0000000
W1:D3-W1:D1 122.159091   86.006418 158.311764 0.0000000
W2:D4-W1:D1 276.337662  251.389948 301.285377 0.0000000
W3:D4-W1:D1 259.777512  232.352937 287.202087 0.0000000
W1:D4-W1:D1 296.266234  269.551592 322.980876 0.0000000
W3:D2-W2:D2 102.646199   73.844114 131.448284 0.0000000
W1:D2-W2:D2  10.660088  -16.229734  37.549910 0.9772520
W2:D3-W2:D2 153.568421  125.515595 181.621248 0.0000000
```

```
W3:D3-W2:D2 146.894737  118.484535 175.304939 0.0000000
W1:D3-W2:D2 167.118421  130.212486 204.024356 0.0000000
W2:D4-W2:D2 321.296992  295.269686 347.324299 0.0000000
W3:D4-W2:D2 304.736842  276.326640 333.147044 0.0000000
W1:D4-W2:D2 341.225564  313.500046 368.951082 0.0000000
W1:D2-W3:D2 -91.986111 -119.289647 -64.682575 0.0000000
W2:D3-W3:D2  50.922222   22.472589  79.371855 0.0000008
W3:D3-W3:D2  44.248538   15.446453  73.050623 0.0000513
W1:D3-W3:D2  64.472222   27.263775 101.680670 0.0000021
W2:D4-W3:D2 218.650794  192.196281 245.105306 0.0000000
W3:D4-W3:D2 202.090643  173.288558 230.892728 0.0000000
W1:D4-W3:D2 238.579365  210.452423 266.706307 0.0000000
W2:D3-W1:D2 142.908333  116.396373 169.420294 0.0000000
W3:D3-W1:D2 136.234649  109.344827 163.124471 0.0000000
W1:D3-W1:D2 156.458333  120.709613 192.207053 0.0000000
W2:D4-W1:D2 310.636905  286.278258 334.995551 0.0000000
W3:D4-W1:D2 294.076754  267.186932 320.966576 0.0000000
W1:D4-W1:D2 330.565476  304.400092 356.730860 0.0000000
W3:D3-W2:D3  -6.673684  -34.726511  21.379142 0.9997455
W1:D3-W2:D3  13.550000  -23.081537  50.181537 0.9867743
W2:D4-W2:D3 167.728571  142.091836 193.365307 0.0000000
W3:D4-W2:D3 151.168421  123.115595 179.221248 0.0000000
W1:D4-W2:D3 187.657143  160.297942 215.016343 0.0000000
W1:D3-W3:D3  20.223684  -16.682250  57.129619 0.8112155
W2:D4-W3:D3 174.402256  148.374949 200.429563 0.0000000
W3:D4-W3:D3 157.842105  129.431904 186.252307 0.0000000
W1:D4-W3:D3 194.330827  166.605309 222.056345 0.0000000
W2:D4-W1:D3 154.178571  119.074025 189.283117 0.0000000
W3:D4-W1:D3 137.618421  100.712486 174.524356 0.0000000
W1:D4-W1:D3 174.107143  137.725653 210.488633 0.0000000
W3:D4-W2:D4 -16.560150  -42.587457   9.467157 0.6224704
W1:D4-W2:D4  19.928571   -5.349591  45.206734 0.2840098
W1:D4-W3:D4  36.488722    8.763204  64.214240 0.0012371
```

From the Tuckey Test we can see that not all parameters of the model are significant and we can see many pairs of combinations that provide a non significant p-value (for example W3:D1-W2:D1).

(i) Using the stepwise method, choose a model based on the AIC criterion starting from the full model (including the main effects and the interaction term). Are all coefficients significant?

Hide

```
weightloss$workout = as.factor(weightloss$workout)
weightloss$diet = as.factor(weightloss$diet)
dummy_variables <- model.matrix(~ workout - 1, data = weightloss)
weightloss <- cbind(weightloss, dummy_variables)
dummy_variables <- model.matrix(~ diet - 1, data = weightloss)
weightloss <- cbind(weightloss, dummy_variables)

fitnull<-lm(loss ~ workoutW1 +workoutW2 + workoutW3 + dietD1 + dietD2 + dietD3 + dietD4 + wor
koutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + workoutW1:dietD4 + workoutW2:dietD1 + w
orkoutW2:dietD2 + workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + workoutW3:dietD2 +
workoutW3:dietD3 + workoutW3:dietD4,data=weightloss)
stepSR<-step(fitnull, scope=list(lower = ~ 1,
                                 upper = ~ workoutW1 +workoutW2 + workoutW3 + dietD1 + dietD2
+ dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + workoutW1:dietD4
+ workoutW2:dietD1 + workoutW2:dietD2 + workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD
1 + workoutW3:dietD2 + workoutW3:dietD3 + workoutW3:dietD4),
             direction="both", data=weightloss)
```

```
Start:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + 
    workoutW3:dietD2 + workoutW3:dietD3 + workoutW3:dietD4


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + 
    workoutW3:dietD2 + workoutW3:dietD3


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + 
    workoutW3:dietD2


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW2:dietD1 + workoutW2:dietD2 + workoutW2:dietD3


Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW2:dietD1 + workoutW2:dietD2 + workoutW2:dietD3
```

```
Step:  AIC=1585.01
loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD1 + 
    workoutW1:dietD2 + workoutW1:dietD3 + workoutW2:dietD1 + 
    workoutW2:dietD2 + workoutW2:dietD3

                   Df Sum of Sq    RSS    AIC
- workoutW2:dietD1  1         0 160305 1583.0
- workoutW1:dietD1  1         4 160309 1583.0
- workoutW2:dietD3  1       512 160817 1583.8
- workoutW1:dietD3  1       952 161257 1584.4
<none>                          160305 1585.0
- workoutW2:dietD2  1     72304 232609 1672.3
- workoutW1:dietD2  1     83585 243890 1683.7

Step:  AIC=1583.01
loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD1 + 
    workoutW1:dietD2 + workoutW1:dietD3 + workoutW2:dietD2 + 
    workoutW2:dietD3

                   Df Sum of Sq    RSS    AIC
- workoutW1:dietD1  1         6 160311 1581.0
- workoutW2:dietD3  1       647 160953 1582.0
- workoutW1:dietD3  1      1001 161306 1582.5
<none>                          160305 1583.0
+ workoutW2:dietD1  1         0 160305 1585.0
- workoutW1:dietD2  1     90363 250668 1688.3
- workoutW2:dietD2  1     91886 252191 1689.8

Step:  AIC=1581.02
loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD2 + 
    workoutW1:dietD3 + workoutW2:dietD2 + workoutW2:dietD3

                   Df Sum of Sq     RSS    AIC
- workoutW2:dietD3  1       656  160967 1580.0
- workoutW1:dietD3  1      1246  161557 1580.9
<none>                           160311 1581.0
+ workoutW1:dietD1  1         6  160305 1583.0
+ workoutW2:dietD1  1         2  160309 1583.0
- workoutW2:dietD2  1     92221  252533 1688.1
- workoutW1:dietD2  1    115582  275894 1709.3
- dietD1            1   2847391 3007703 2282.7

Step:  AIC=1580
loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD2 + 
    workoutW1:dietD3 + workoutW2:dietD2

                   Df Sum of Sq     RSS    AIC
- workoutW1:dietD3  1       733  161701 1579.1
<none>                           160967 1580.0
+ workoutW2:dietD3  1       656  160311 1581.0
+ workoutW2:dietD1  1        76  160891 1581.9
+ workoutW1:dietD1  1        15  160953 1582.0
- workoutW2:dietD2  1     96430  257397 1690.7
- workoutW1:dietD2  1    115793  276760 1708.1
```

```
- dietD1            1   2870636 3031603 2282.6

Step:  AIC=1579.09
loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD2 + 
    workoutW2:dietD2

                   Df Sum of Sq     RSS    AIC
<none>                           161701 1579.1
+ workoutW1:dietD3  1       733  160967 1580.0
+ workoutW1:dietD1  1       204  161497 1580.8
+ workoutW2:dietD3  1       144  161557 1580.9
+ workoutW2:dietD1  1        17  161684 1581.1
- workoutW2:dietD2  1     96431  258132 1689.3
- workoutW1:dietD2  1    117266  278967 1708.0
- dietD3            1    745403  907104 1991.0
- dietD1            1   2869903 3031603 2280.6
```

Hide

```
final_model = lm(loss~workoutW1 +workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD2 + wo
rkoutW2:dietD2, data=weightloss )
summary(final_model)
```

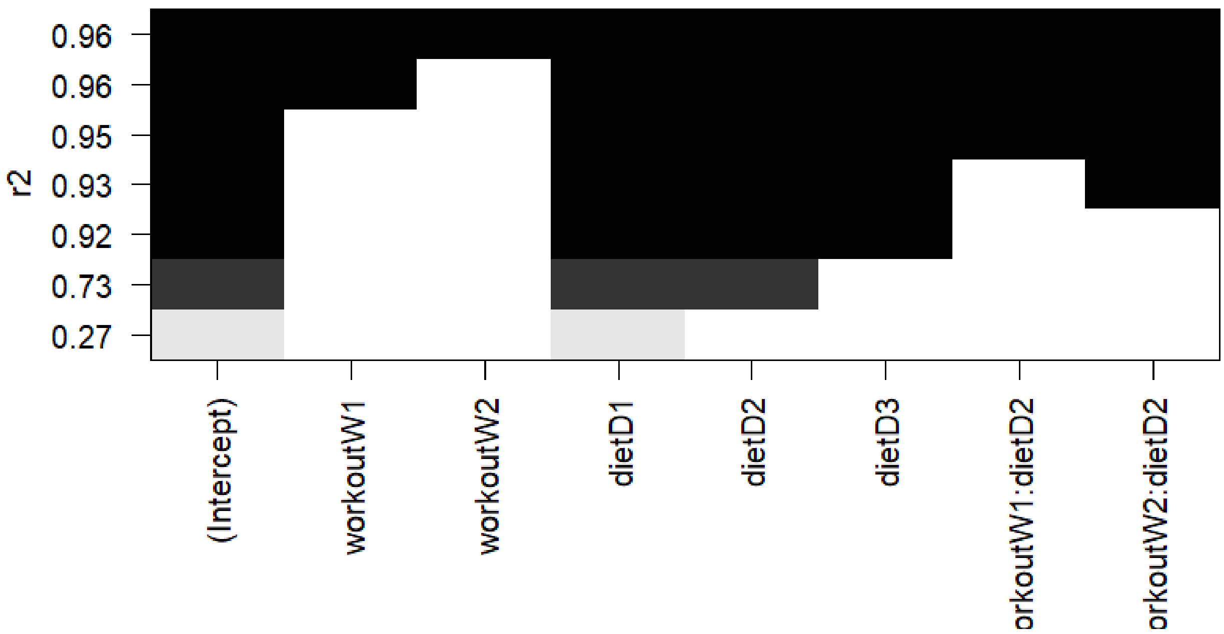
```
Call:
lm(formula = loss ~ workoutW1 + workoutW2 + dietD1 + dietD2 + 
    dietD3 + workoutW1:dietD2 + workoutW2:dietD2, data = weightloss)

Residuals:
    Min      1Q  Median      3Q     Max 
-65.910 -17.769  -1.066  18.106  90.009 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       590.645      4.393 134.462  < 2e-16 ***
workoutW1          33.201      5.038   6.590 2.93e-10 ***
workoutW2          13.496      4.700   2.871  0.00446 ** 
dietD1           -297.151      4.631 -64.169  < 2e-16 ***
dietD2           -204.368      7.617 -26.831  < 2e-16 ***
dietD3           -165.231      5.053 -32.703  < 2e-16 ***
workoutW1:dietD2 -125.187      9.651 -12.971  < 2e-16 ***
workoutW2:dietD2 -116.142      9.874 -11.762  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.4 on 232 degrees of freedom
Multiple R-squared:  0.9607,    Adjusted R-squared:  0.9595 
F-statistic: 810.3 on 7 and 232 DF,  p-value: < 2.2e-16
```

Yes all the coefficients of the final model are statistically significant.

(j) Provide a graphical representation for the final model.

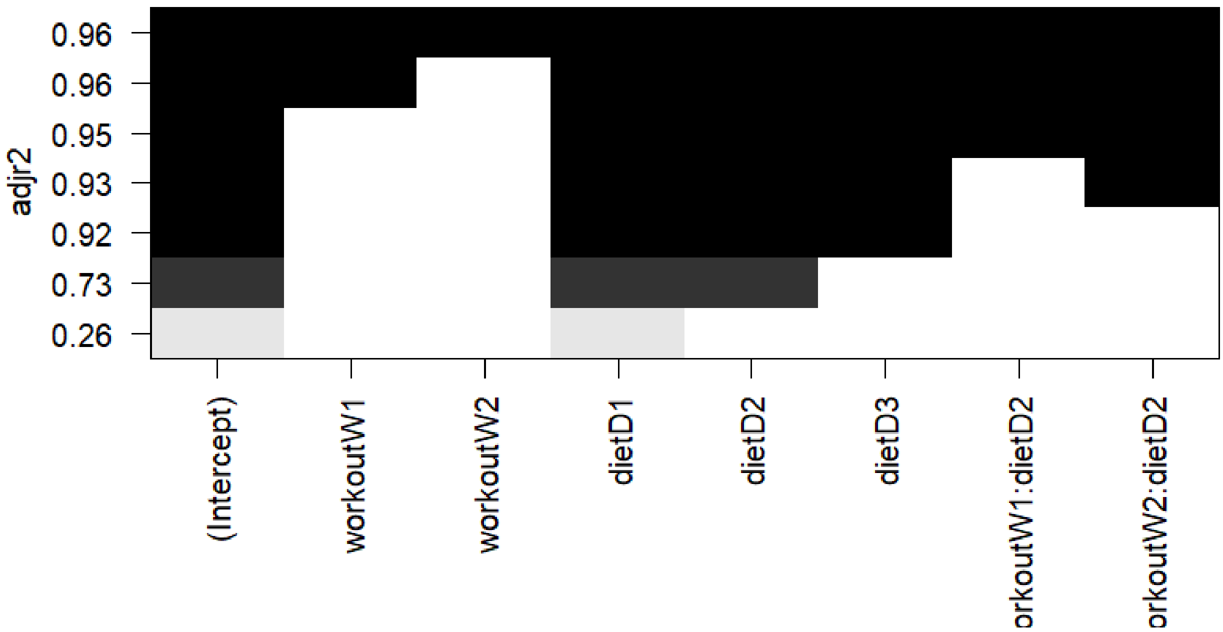
Hide

```
leaps<-regsubsets(loss~workoutW1 +workoutW2 + dietD1 + dietD2 + dietD3 + workoutW1:dietD2 + w
orkoutW2:dietD2, data=weightloss )
plot(leaps,scale="r2")
```



Hide

```
plot(leaps,scale="adjr2")
```



(k) Compare the constant model against the (full) main effects model and the (full) interaction model. Are the models different?

Hide

```
fitM0<-lm(loss~workoutW1 +workoutW2 + workoutW3 + dietD1 + dietD2 + dietD3 + dietD4 + workout
W1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + workoutW1:dietD4 + workoutW2:dietD1 + worko
utW2:dietD2 + workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + workoutW3:dietD2 + wor
koutW3:dietD3 + workoutW3:dietD4,data=weightloss)
fitM1<-lm(loss~workoutW1 +workoutW2 + workoutW3 + dietD1 + dietD2 + dietD3 + dietD4,data=weig
htloss)
fitM2<-lm(loss~1,data=weightloss)
anova(fitM0,fitM2)
```

```
Analysis of Variance Table

Model 1: loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 + 
    dietD3 + dietD4 + workoutW1:dietD1 + workoutW1:dietD2 + workoutW1:dietD3 + 
    workoutW1:dietD4 + workoutW2:dietD1 + workoutW2:dietD2 + 
    workoutW2:dietD3 + workoutW2:dietD4 + workoutW3:dietD1 + 
    workoutW3:dietD2 + workoutW3:dietD3 + workoutW3:dietD4
Model 2: loss ~ 1
  Res.Df     RSS  Df Sum of Sq      F    Pr(>F)    
1    228  160305                                   
2    239 4115147 -11  -3954842 511.36 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
anova(fitM1,fitM2)
```

```
Analysis of Variance Table

Model 1: loss ~ workoutW1 + workoutW2 + workoutW3 + dietD1 + dietD2 +
    dietD3 + dietD4
Model 2: loss ~ 1
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    234  302557
2    239 4115147 -5  -3812590 589.74 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
aic_model0 <- AIC(fitM0)
aic_model1 <- AIC(fitM1)
aic_model2 <- AIC(fitM2)
print(c(AIC = c(aic_model0,aic_model1, aic_model2)))
```

```
    AIC1     AIC2     AIC3
2268.097 2408.543 3024.982
```

Yes, the models are completely different. It can be seen from the AIC values of the larger models that are much lower from the simple model AIC, even though they have more params, meaning that they have a significantly larger likelihood value.