5.1\_lab.R

vladv

2024-11-02

#Дисперсионный анализ. Лабораторная  
#Малявко Е.С.  
#РИ-411055  
install.packages("gplots")

## Error in install.packages : Updating loaded packages

install.packages("multcomp")

## Error in install.packages : Updating loaded packages

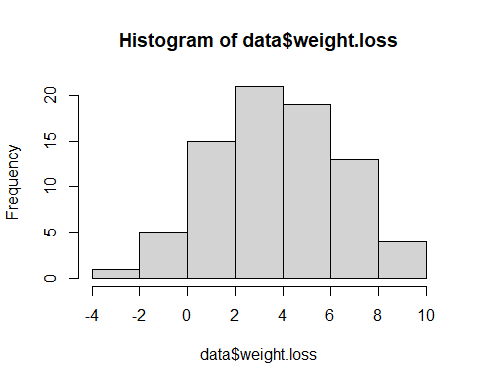
#Загрузим данные (требуется установить Рабочую папку с помощью setwd) или указать полный путь  
data = read.csv("C:/Users/vladv/Downloads/lecture 5.1/data/diet.csv",row.names=1)  
summary(data)

## gender Age Height   
## Min. :0.0000 Min. :16.00 Min. :141.0   
## 1st Qu.:0.0000 1st Qu.:32.25 1st Qu.:164.2   
## Median :0.0000 Median :39.00 Median :169.5   
## Mean :0.4342 Mean :39.15 Mean :170.8   
## 3rd Qu.:1.0000 3rd Qu.:46.75 3rd Qu.:174.8   
## Max. :1.0000 Max. :60.00 Max. :201.0   
## NA's :2   
## pre.weight Diet weight6weeks   
## Min. : 58.00 Min. :1.000 Min. : 53.00   
## 1st Qu.: 66.00 1st Qu.:1.000 1st Qu.: 61.85   
## Median : 72.00 Median :2.000 Median : 68.95   
## Mean : 72.53 Mean :2.038 Mean : 68.68   
## 3rd Qu.: 78.00 3rd Qu.:3.000 3rd Qu.: 73.83   
## Max. :103.00 Max. :3.000 Max. :103.00   
##

#Ознакомимся со структурой и переименуем колонки, как нам удобно  
#files/Diet\_data\_description.docx  
  
colnames(data) <- c("gender", "age", "height", "initial.weight",   
 "diet.type", "final.weight")  
data$diet.type <- factor(c("A", "B", "C")[data$diet.type])  
#Добавим новую колонку - Похудение  
data$weight.loss = data$initial.weight - data$final.weight  
summary(data)

## gender age height   
## Min. :0.0000 Min. :16.00 Min. :141.0   
## 1st Qu.:0.0000 1st Qu.:32.25 1st Qu.:164.2   
## Median :0.0000 Median :39.00 Median :169.5   
## Mean :0.4342 Mean :39.15 Mean :170.8   
## 3rd Qu.:1.0000 3rd Qu.:46.75 3rd Qu.:174.8   
## Max. :1.0000 Max. :60.00 Max. :201.0   
## NA's :2   
## initial.weight diet.type final.weight weight.loss   
## Min. : 58.00 A:24 Min. : 53.00 Min. :-2.100   
## 1st Qu.: 66.00 B:27 1st Qu.: 61.85 1st Qu.: 2.000   
## Median : 72.00 C:27 Median : 68.95 Median : 3.600   
## Mean : 72.53 Mean : 68.68 Mean : 3.845   
## 3rd Qu.: 78.00 3rd Qu.: 73.83 3rd Qu.: 5.550   
## Max. :103.00 Max. :103.00 Max. : 9.200   
##

hist(data$weight.loss) #на общей гистограмме выбросов не видно, разделим на группы по виду диеты



data.A <- subset(data, diet.type == "A")  
data.B <- subset(data, diet.type == "B")  
data.C <- subset(data, diet.type == "C")  
  
out.A <- boxplot.stats(data.A$weight.loss)$out #выбросы по каждой группе  
out.B <- boxplot.stats(data.B$weight.loss)$out  
out.C <- boxplot.stats(data.C$weight.loss)$out  
  
out.A

## [1] 8.5 9.0

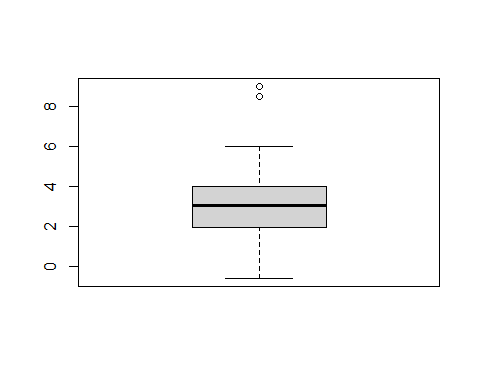
out.B

## numeric(0)

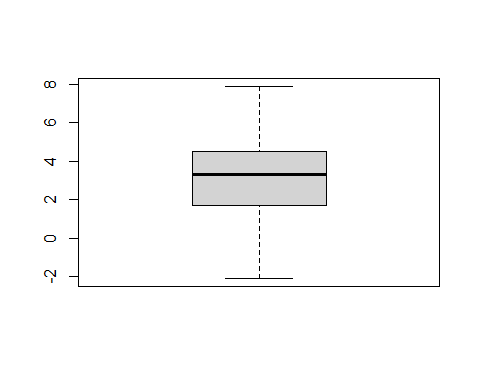
out.C

## numeric(0)

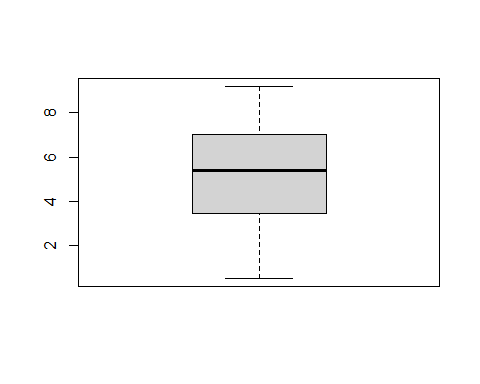
boxplot(data.A$weight.loss) #на диаграмме видны выбросы в А-группе



boxplot(data.B$weight.loss) #здесь выбросов нет



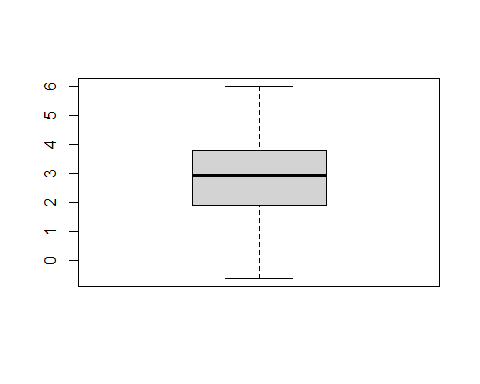
boxplot(data.C$weight.loss) #здесь выбросов нет



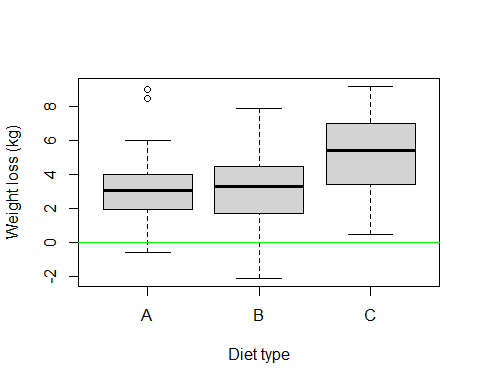
#выбросы есть только в А-группе, удалим их:  
  
fixeddata <- subset(data, weight.loss != 8.5 & weight.loss != 9.0)  
fixeddata.A <- subset(fixeddata, diet.type == "A")  
boxplot.stats(fixeddata.A$weight.loss)$out# теперь выбросов нет

## numeric(0)

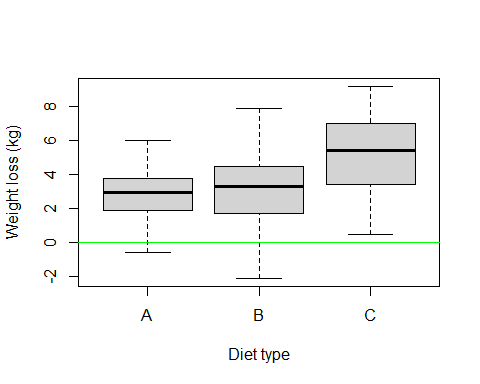
boxplot(fixeddata.A$weight.loss)



boxplot(fixeddata.A$weight.loss) #диаграмма А-группы без выбросов  
  
  
#Проанализиуем есть ли различия по типам диет  
  
#с выбросами:  
boxplot(weight.loss~diet.type,data=data,col="light gray",  
 ylab = "Weight loss (kg)", xlab = "Diet type")  
abline(h=0,col="green")



#без выбросов:  
boxplot(weight.loss~diet.type,data=fixeddata,col="light gray",  
 ylab = "Weight loss (kg)", xlab = "Diet type")  
abline(h=0,col="green")



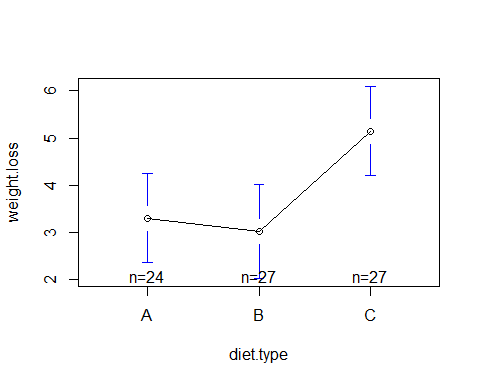
#без выбросов потеря веса при А-диете меньше, чем с выбросами  
  
#проверим сбалансированные ли данные  
table(data$diet.type)

##   
## A B C   
## 24 27 27

table(fixeddata$diet.type)

##   
## A B C   
## 22 27 27

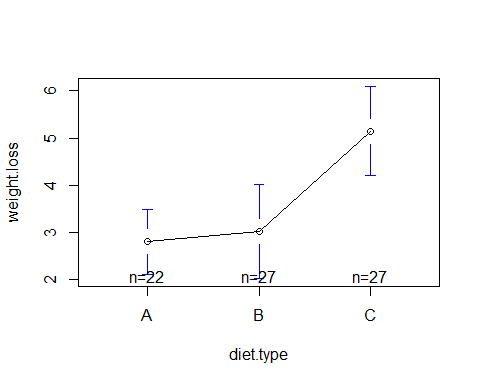
#График групповых средних  
#с выбросами:  
library(gplots) #библиотека устанавлевается с помощью install.packages  
plotmeans(weight.loss ~ diet.type, data=data)



aggregate(data$weight.loss, by = list(data$diet.type), FUN=sd)

## Group.1 x  
## 1 A 2.240148  
## 2 B 2.523367  
## 3 C 2.395568

#без выбросов:  
library(gplots) #библиотека устанавлевается с помощью install.packages  
plotmeans(weight.loss ~ diet.type, data=fixeddata)



aggregate(fixeddata$weight.loss, by = list(fixeddata$diet.type), FUN=sd)

## Group.1 x  
## 1 A 1.550569  
## 2 B 2.523367  
## 3 C 2.395568

#без выбросов среднеквадратичное отклонение меньше, на графике   
#групповых средних это также видно  
  
#Для подгонки ANOVA модели используем функцию aov, частный случай линейной модели lm  
#с выбросами:  
fit <- aov(weight.loss ~ diet.type, data=data)  
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)   
## diet.type 2 71.1 35.55 6.197 0.00323 \*\*  
## Residuals 75 430.2 5.74   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#без выбросов:  
fixed.fit <- aov(weight.loss ~ diet.type, data=fixeddata)  
summary(fixed.fit)

## Df Sum Sq Mean Sq F value Pr(>F)   
## diet.type 2 86.5 43.26 8.645 0.000427 \*\*\*  
## Residuals 73 365.2 5.00   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

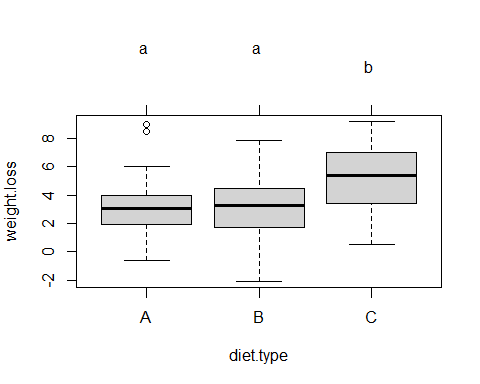
#попарные различия между средними значениями для всех групп  
#с выбросами:  
TukeyHSD(fit)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = weight.loss ~ diet.type, data = data)  
##   
## $diet.type  
## diff lwr upr p adj  
## B-A -0.2740741 -1.8806155 1.332467 0.9124737  
## C-A 1.8481481 0.2416067 3.454690 0.0201413  
## C-B 2.1222222 0.5636481 3.680796 0.0047819

#без выбросов:  
TukeyHSD(fixed.fit)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = weight.loss ~ diet.type, data = fixeddata)  
##   
## $diet.type  
## diff lwr upr p adj  
## B-A 0.2213805 -1.3156340 1.758395 0.9367018  
## C-A 2.3436027 0.8065882 3.880617 0.0014162  
## C-B 2.1222222 0.6657364 3.578708 0.0023769

#Tukey honest significant differences test  
#с выбросами:  
library(multcomp)  
par(mar=c(5,4,6,2))  
tuk <- glht(fit, linfct=mcp(diet.type="Tukey"))  
plot(cld(tuk, level=.05),col="lightgrey")



#без выбросов:  
library(multcomp)  
par(mar=c(5,4,6,2))  
tuk <- glht(fixed.fit, linfct=mcp(diet.type="Tukey"))  
plot(cld(tuk, level=.05),col="lightgrey")

