Lab 5.1

Mineev Anatolii

2024-10-28

One-way ANOVA

```
data = read.csv("../data/diet.csv",row.names=1)
summary(data)
```

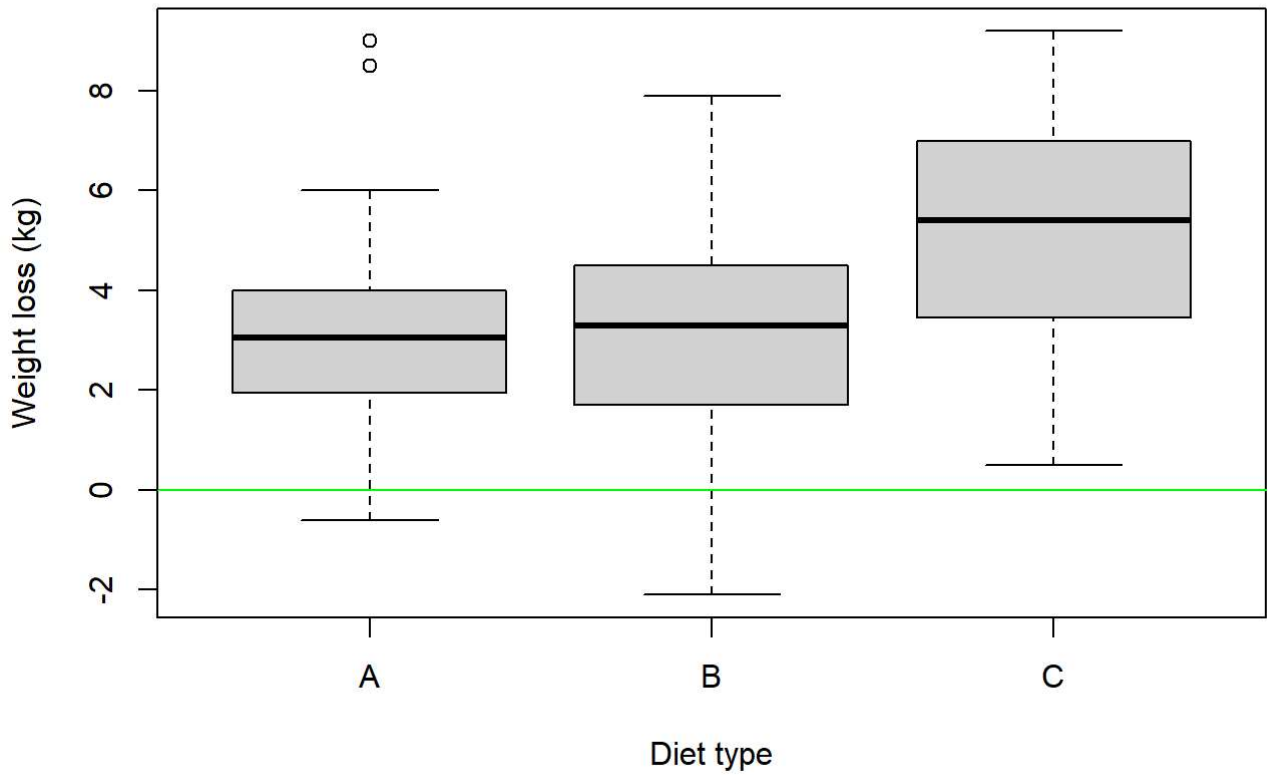
```
##      gender            Age            Height        pre.weight
##  Min.   :0.0000   Min.   :16.00   Min.   :141.0   Min.   : 58.00
##  1st Qu.:0.0000   1st Qu.:32.25   1st Qu.:164.2   1st Qu.: 66.00
##  Median :0.0000   Median :39.00   Median :169.5   Median : 72.00
##  Mean   :0.4342   Mean 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

Обработка данных:

```
colnames(data) <- c("gender", "age", "height", "initial.weight",
                    "diet.type", "final.weight")
data$diet.type <- factor(c("A", "B", "C")[data$diet.type])
data$gender <- factor(c("F", "M")[data$gender + 1])
#Добавим новую колонку - Похудение
data$weight.loss = data$initial.weight - data$final.weight
summary(data)
```

```
##   gender        age            height      initial.weight   diet.type
##  F   :43   Min.   :16.00   Min.   :141.0   Min.   : 58.00   A:24
##  M   :33   1st Qu.:32.25   1st Qu.:164.2   1st Qu.: 66.00   B:27
##  NA's: 2   Median :39.00   Median :169.5   Median : 72.00   C:27
##            Mean   :39.15   Mean   :170.8   Mean   : 72.53
##            3rd Qu.:46.75   3rd Qu.:174.8   3rd Qu.: 78.00
##            Max.   :60.00   Max.   :201.0   Max.   :103.00
##   final.weight     weight.loss
##  Min.   : 53.00   Min.   :-2.100
##  1st Qu.: 61.85   1st Qu.: 2.000
##  Median : 68.95   Median : 3.600
##  Mean   : 68.68   Mean   : 3.845
##  3rd Qu.: 73.83   3rd Qu.: 5.550
##  Max.   :103.00   Max.   : 9.200
```

```
#Проанализируем есть ли различия по типам диет
boxplot(weight.loss~diet.type,data=data,col="light gray",
        ylab = "Weight loss (kg)", xlab = "Diet type")
abline(h=0,col="green")
```




```
#проверим сбалансированные ли данные
table(data$diet.type)
```

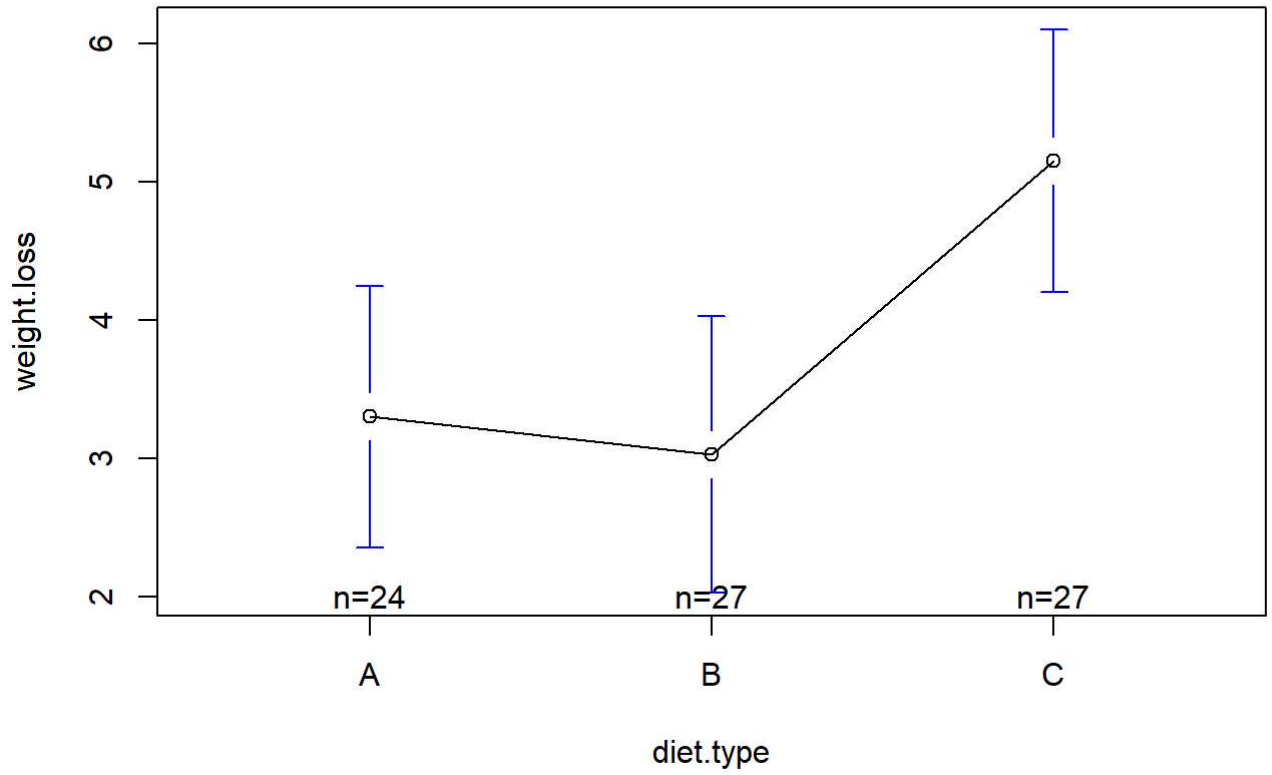
```
##
##  A  B  C
## 24 27 27
```

```
#График групповых средних
library(gplots) #библиотека устанавлевается с помощью install.packages
```

```
##
## Присоединяю пакет: 'gplots'
```

```
## Следующий объект скрыт от 'package:stats':
##
##     lowess
```

```
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
```

```
#Для подгонки ANOVA модели используем функцию aov, частный случай линейной модели lm
#тест на межгрупповые различия
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
```

```
#попарные различия между средними значениями для всех групп
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
```

```
#Tukey honest significant differences test)
library(multcomp)
```

```
## Загрузка требуемого пакета: mvtnorm
```

```
## Загрузка требуемого пакета: survival
```

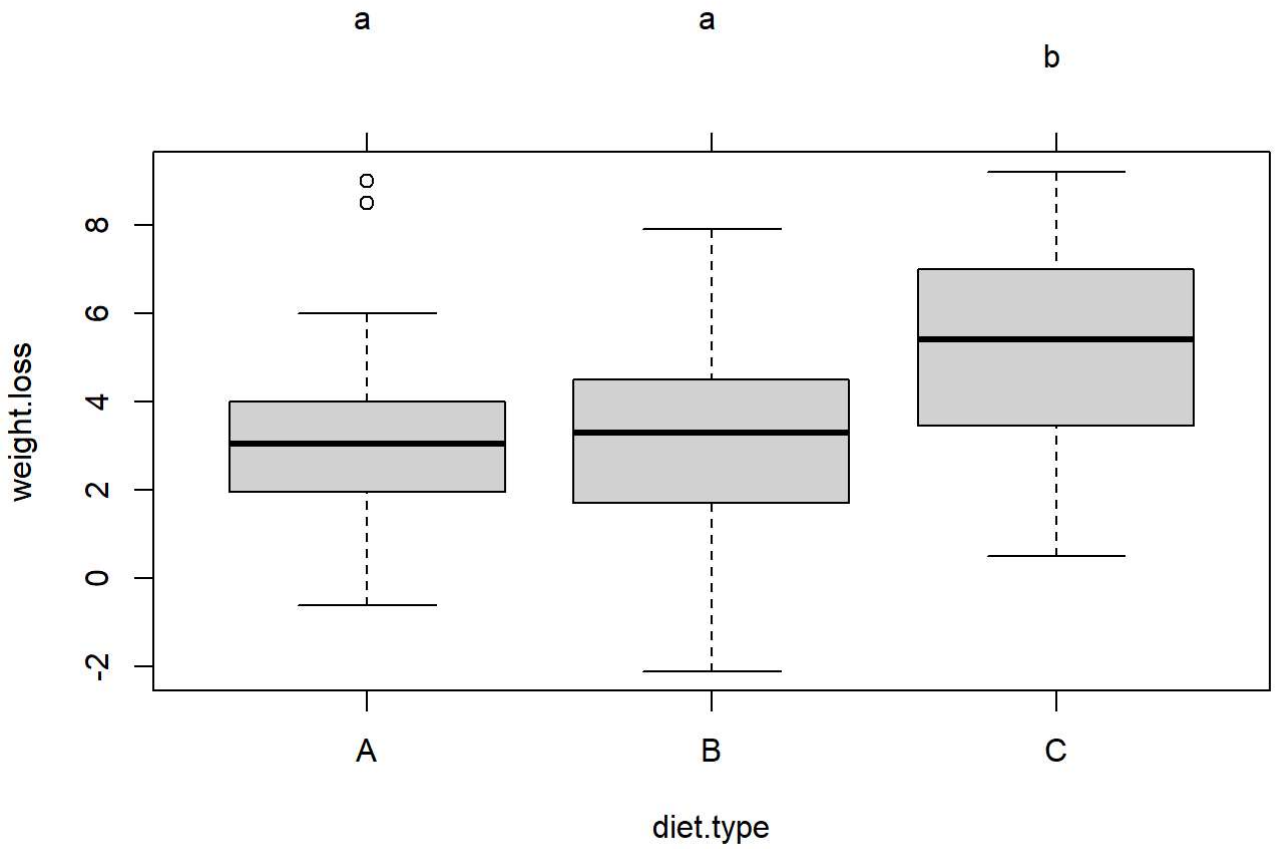
```
## Загрузка требуемого пакета: TH.data
```

```
## Загрузка требуемого пакета: MASS
```

```
##
## Присоединяю пакет: 'TH.data'
```

```
## Следующий объект скрыт от 'package:MASS':
##
##     geyser
```

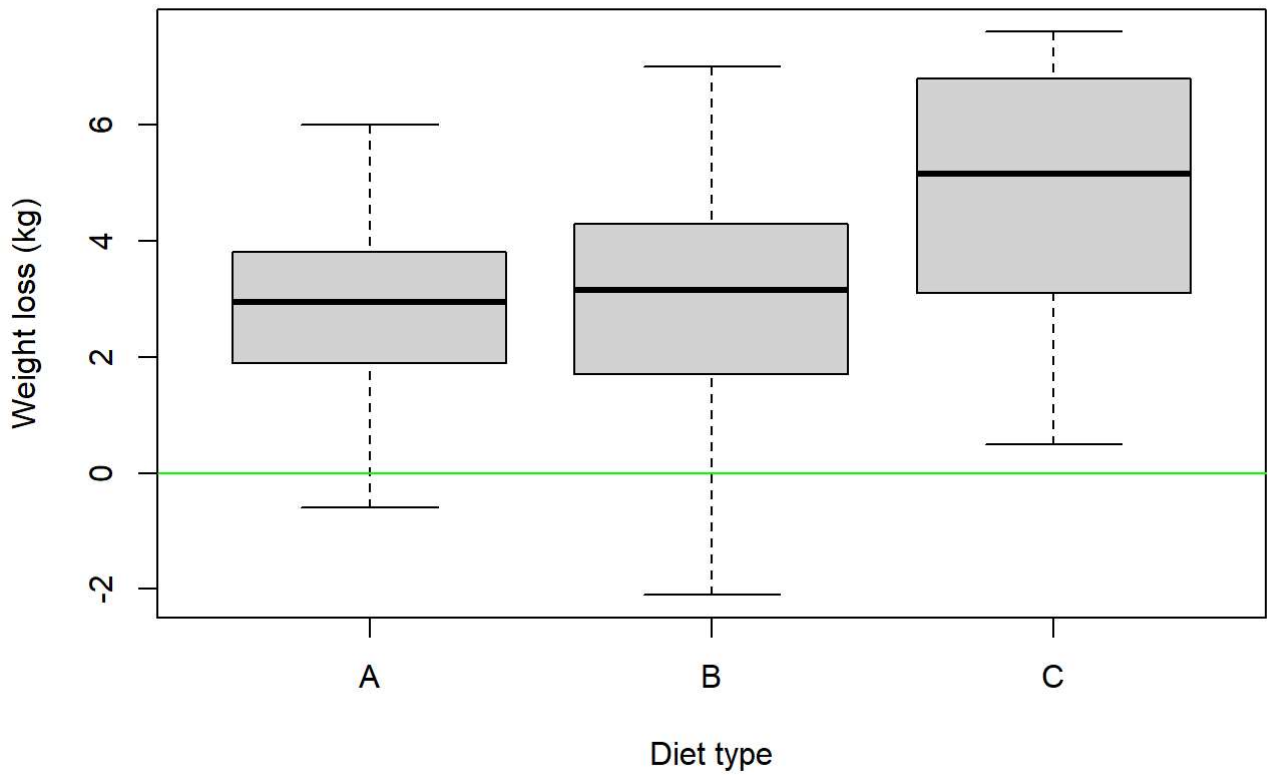
```
par(mar=c(5,4,6,2))
tuk <- glht(fit, linfct=mcp(diet.type="Tukey"))
plot(cld(tuk, level=.05),col="lightgrey")
```



One-way ANOVA без выбросов

Удалим выбросы (данные с z-score > 3):

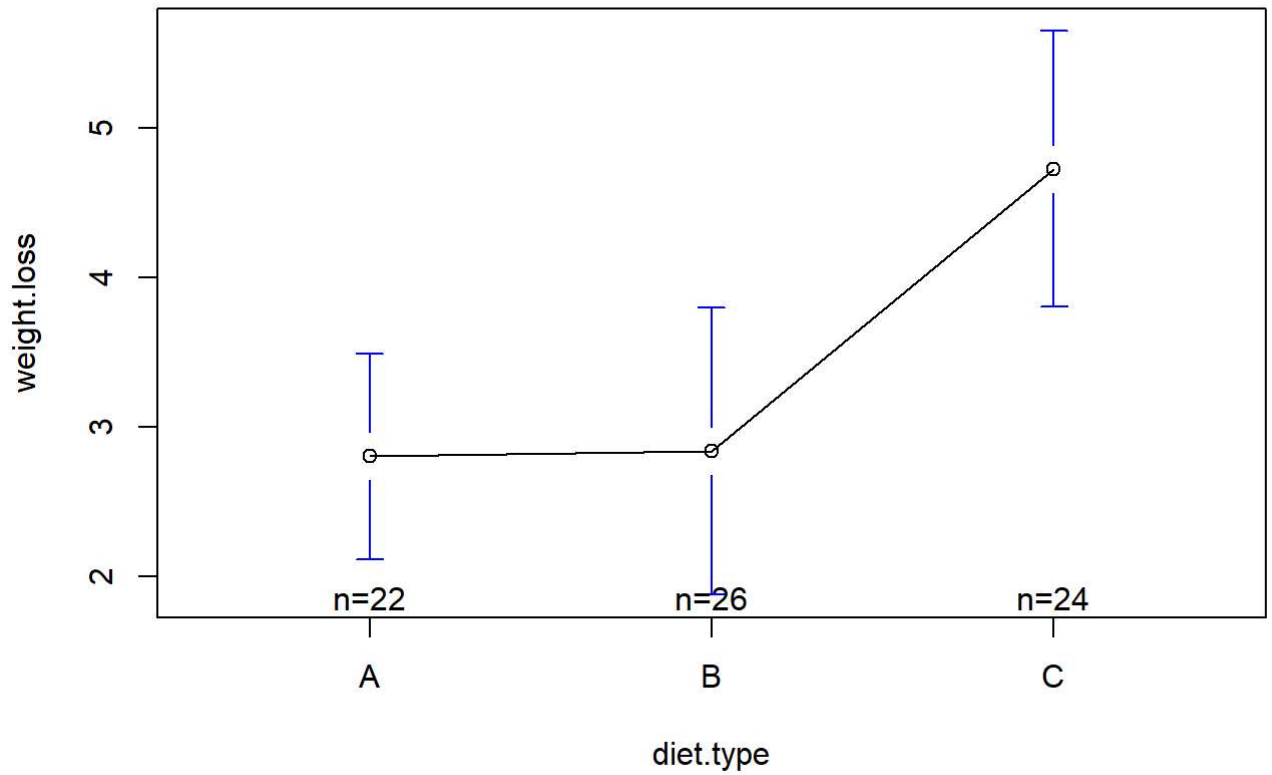
```
sko <- sd(data$weight.loss)
d1 <- data$weight.loss

d2 <- d1[d1 > 3.5 * sko]


fixed.data <- data[data$weight.loss <= 3 * sko, ]

boxplot(weight.loss~diet.type,data=fixed.data,col="light gray",
        ylab = "Weight loss (kg)", xlab = "Diet type")
abline(h=0,col="green")
```



```
#проверим сбалансированные ли данные
table(fixed.data$diet.type)
```

```
##
##  A  B  C
## 22 26 24
```

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




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

```
##   Group.1        x
## 1       A 1.550569
## 2       B 2.373871
## 3       C 2.182390
```

```
#Для подгонки ANOVA модели используем функцию aov, частный случай линейной модели lm
#тест на межгрупповые различия
fit <- aov(weight.loss ~ diet.type, data=fixed.data)
summary(fit)
```

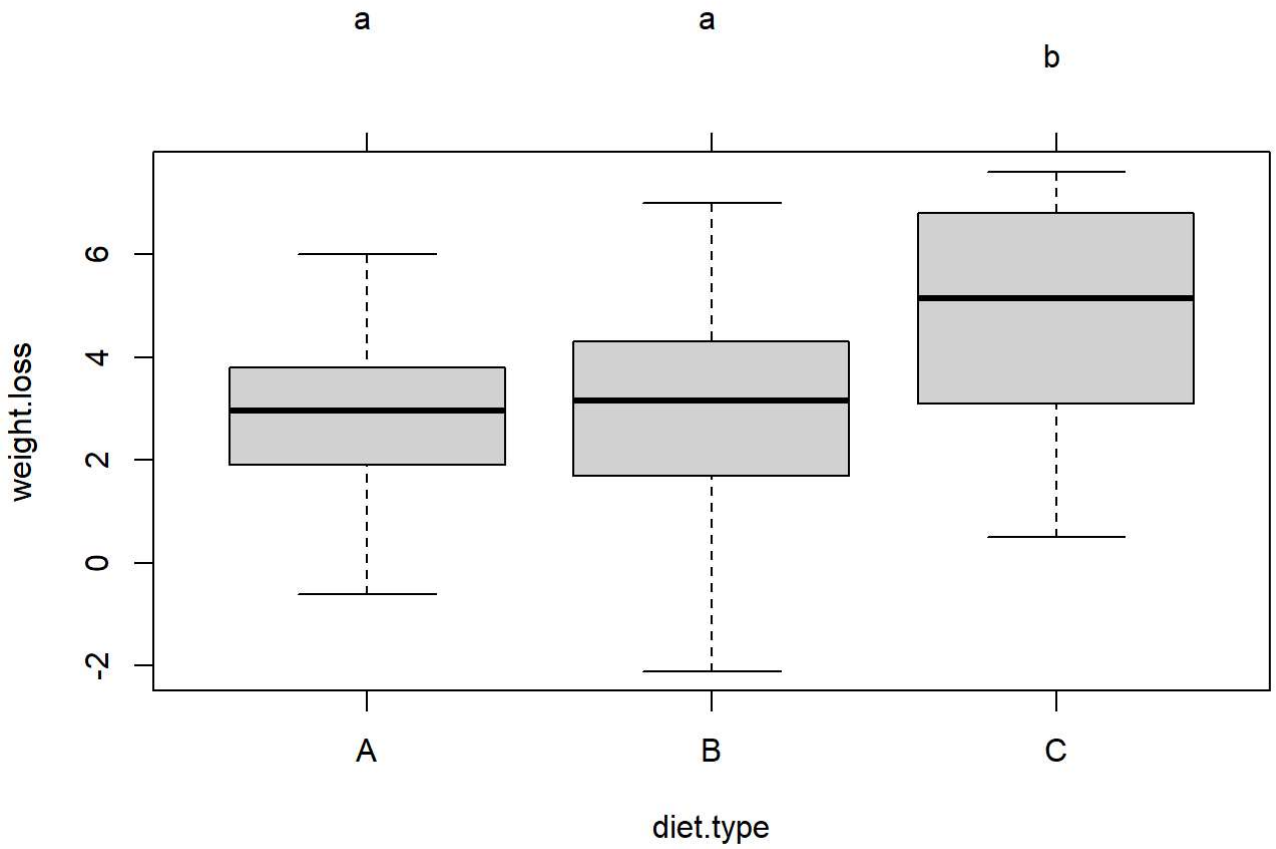
```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## diet.type    2   57.9  28.950   6.638 0.00231 **
## Residuals   69  300.9   4.361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

В этот раз p-value меньше

```
TukeyHSD(fit)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight.loss ~ diet.type, data = fixed.data)
##
## $diet.type
##           diff        lwr      upr     p adj
## B-A 0.03391608 -1.4151321 1.482964 0.9982686
## C-A 1.92045455  0.4439928 3.396916 0.0074471
## C-B 1.88653846  0.4705721 3.302505 0.0059695
```

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



Two-way ANOVA

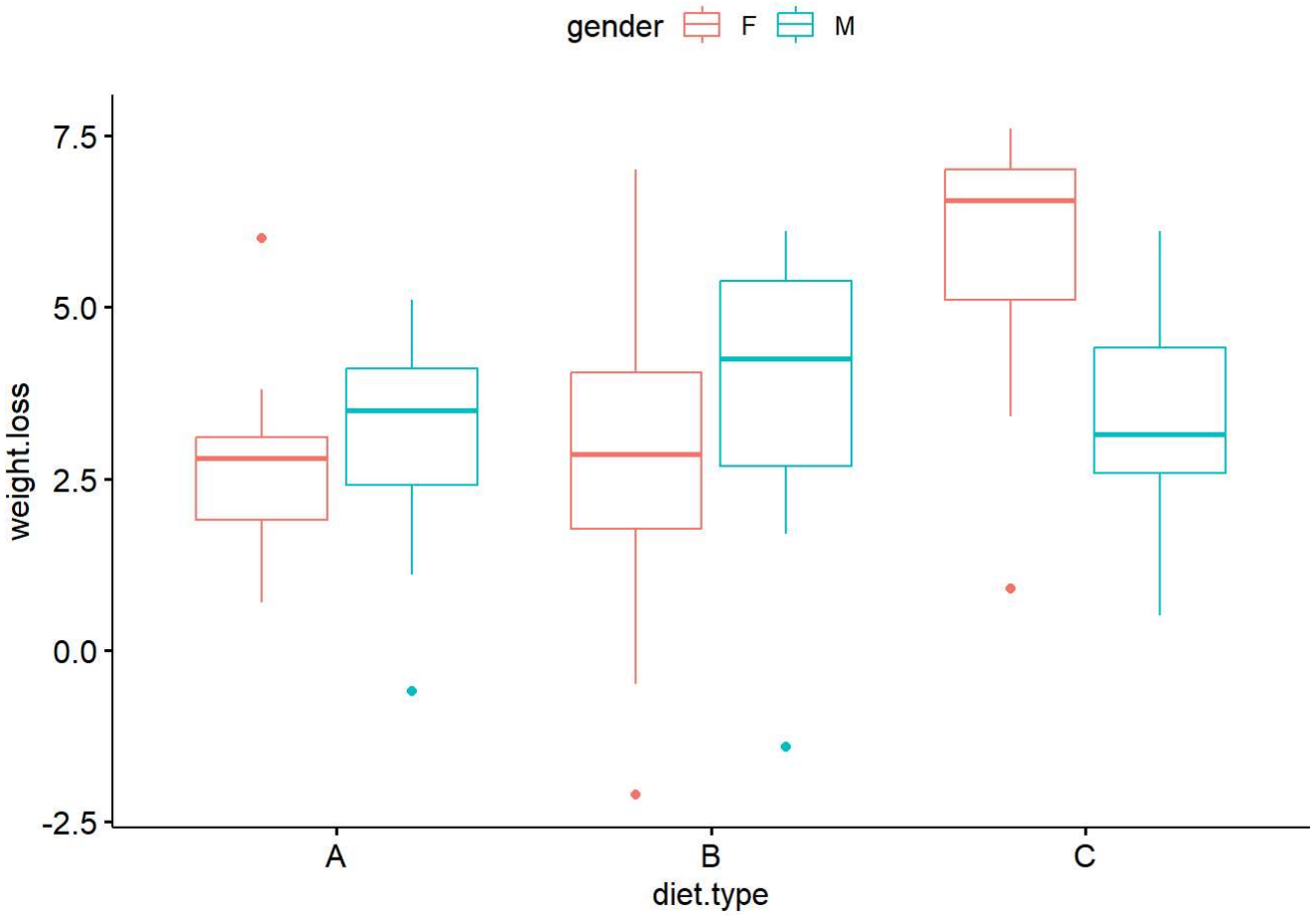
```
gender.data <- fixed.data[complete.cases(fixed.data), ]
table(gender.data$gender, gender.data$diet.type)
```

```
##
##      A  B  C
##   F 13 14 14
##   M  9 10 10
```

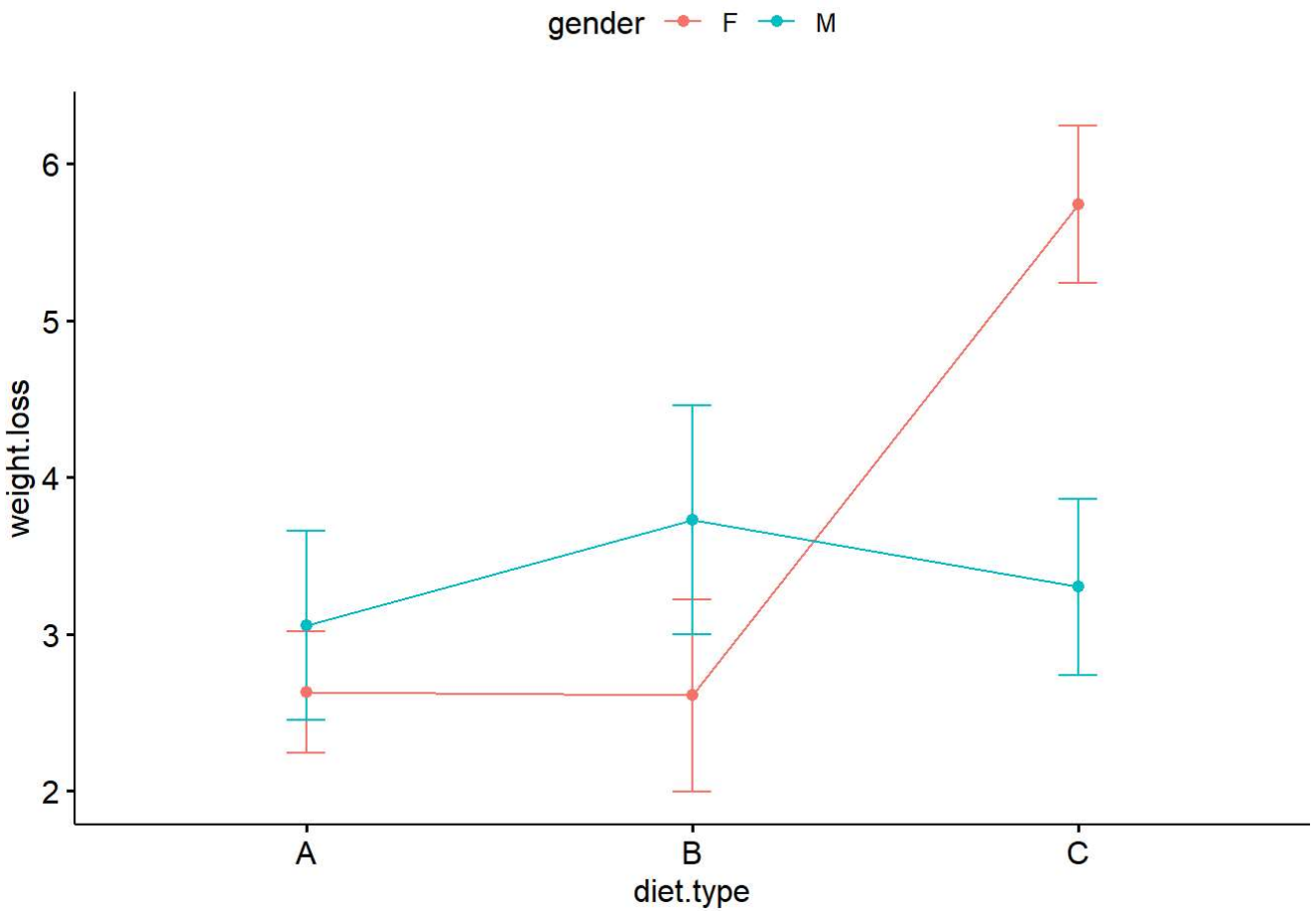
```
library("ggpubr")
```

```
## Загрузка требуемого пакета: ggplot2
```

```
ggboxplot(gender.data, x="diet.type", y="weight.loss", color="gender")
```




```
ggline(gender.data, x="diet.type", y="weight.loss", color="gender", add = c("mean_se"))
```




```
fit.gender <- aov(weight.loss ~ diet.type*gender, data=gender.data)
summary(fit.gender)
```

```
##                  Df Sum Sq Mean Sq F value  Pr(>F)
## diet.type         2  50.77  25.387   6.760 0.00217 **
## gender            1   1.74   1.743   0.464 0.49816
## diet.type:gender  2  41.38  20.691   5.510 0.00620 **
## Residuals        64 240.33   3.755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANCOVA

```
fit.height <- aov(weight.loss ~ diet.type*gender*height, data=gender.data)
summary(fit.height)
```

```
##                          Df Sum Sq Mean Sq F value  Pr(>F)
## diet.type                 2  50.77  25.387   6.555 0.00271 **
## gender                    1   1.74   1.743   0.450 0.50499
## height                    1   2.60   2.599   0.671 0.41607
## diet.type:gender          2  41.44  20.722   5.350 0.00737 **
## diet.type:height          2  10.11   5.056   1.306 0.27888
## gender:height             1   1.99   1.986   0.513 0.47676
## diet.type:gender:height   2   0.93   0.467   0.121 0.88654
## Residuals                58 224.64   3.873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Рост не влияет на сброшенный вес