

# Analiza wariancji

## Zbiór danych

Zbiór danych zawiera informacje o wychodowanych myszach. Celem badania było określenie wpływu pewnych genów na rozwój symptomów zespołu Downa. Myszy zostały urodzone przez genetycznie zmodyfikowane osobniki (brane były pod uwagę cztery fragmenty kodu DNA), przy czym nie wszystkie odziedziczyły zmodyfikowane geny.

```
mouse <- read.csv("mouse.csv")
```

### Zmienne

zmienna	opis
DNA	Modyfikacja genetyczna rodzica: 1=141G6; 2=152F7; 3=230E8; 4=285E6
line	Linia rodzinna
tg	0 = mysz nie jest transgeniczna, 1 = mysz jest transgeniczna
sex	płeć, 1 = samiec, 0 = samica
age	wiek (dni)
weight	waga (gramy)
cage	numer klatki

## Problemy

1. Wykonaj czyszczenie danych, tak aby były one prawidłowe pod względem formalnym.

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
mouse <- mouse %>% mutate(sex = factor(sex, levels = c(`samica` = 0, samiec = 1), labels = c("samica", "samiec"))) %>% mutate(tg = factor(tg, levels = c(`nie jest transgeniczna` = 0, transgeniczna = 1), labels = c("nie jest transgeniczna", "transgeniczna")))
```

2. Utwórz nową zmienną, która wskazuje czy mysz jest transgeniczna, a jeśli jest to zawiera informację o odziedziczonym fragmencie DNA.

```

mouse['trans'] <- mouse$DNA
mouse$trans = replace(mouse$trans,mouse$tg==1, "nie jest transgeniczna")
mouse$trans = replace(mouse$trans,mouse$trans==1, "141G6")
mouse$trans = replace(mouse$trans,mouse$trans==2, "152F7")
mouse$trans = replace(mouse$trans,mouse$trans==3, "230E8")
mouse$trans = replace(mouse$trans,mouse$trans==4, "285E6")
mouse = mouse %>% mutate (trans = factor(trans))

```

3. Zrób histogram masy ciała myszy. Opisz histogram i zbadaj skąd wynika jego kształt

```

hist(mouse$weight,main = paste("Histogram masy ciała myszy"),ylab = "masa w gramach",xlab = "mysz",xlim = c(20,45))

```

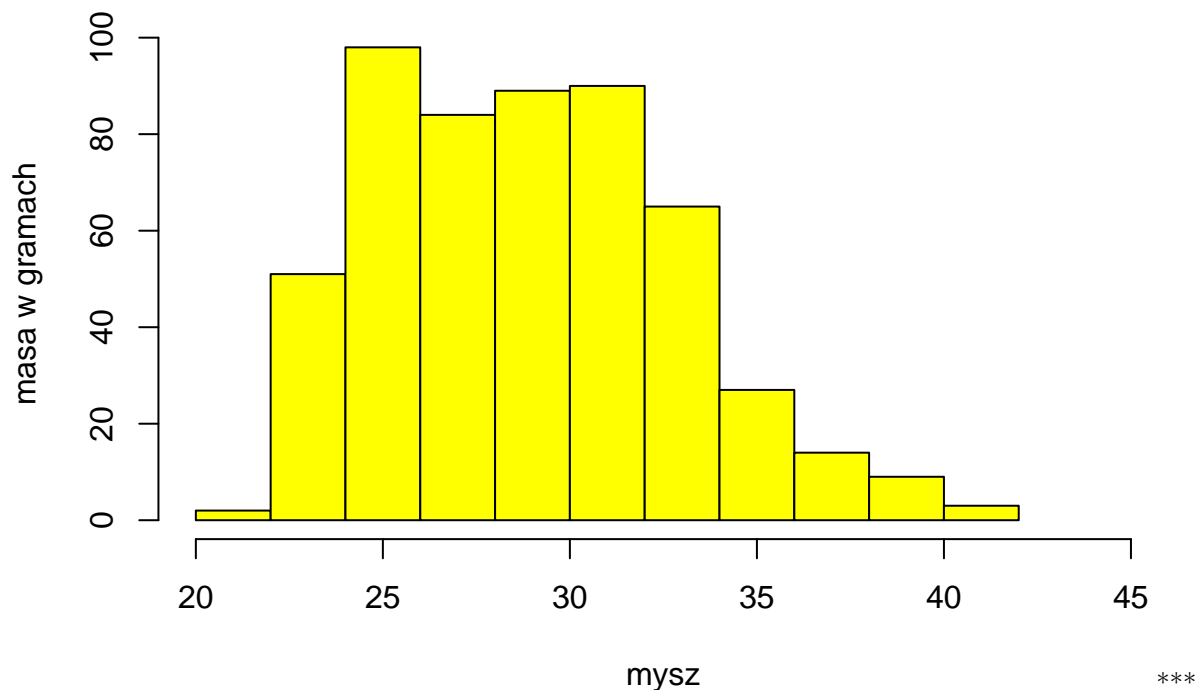
```

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## conversion failure on 'Histogram masy ciała myszy' in 'mbcsToSbcs': dot
## substituted for <c5>

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):
## conversion failure on 'Histogram masy ciała myszy' in 'mbcsToSbcs': dot
## substituted for <82>

```

### Histogram masy ciała myszy



4. Czy ma sens badanie wpływu genów w całej populacji, czy może warto podzielić populację myszy na podgrupy?

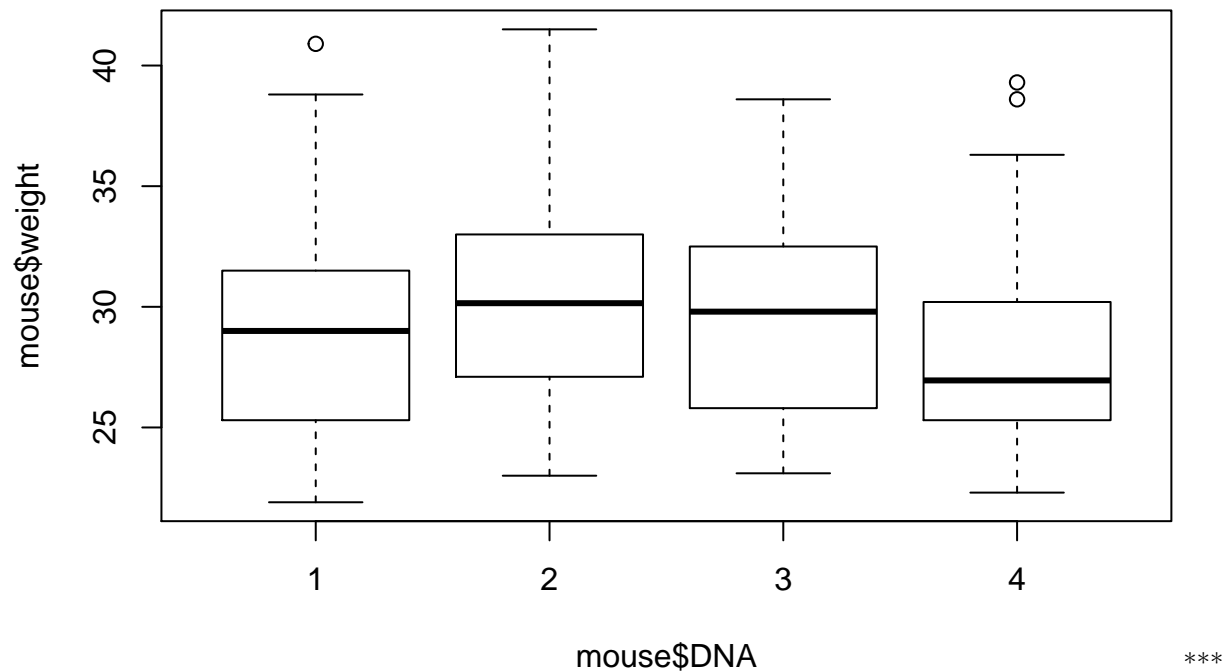
Warto podzielić populację myszy na podgrupy, zobaczymy wtedy więcej zależności. \*\*\*

5. Porównaj myszy transgeniczne z tymi, które odziedziczyły poszczególne geny. Czy zawsze dodatkowe geny powodują wzrost masy ciała myszy?

```
library(gridExtra)

##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##   combine
library(dplyr)

boxplot(mouse$weight~mouse$DNA)
```



6. Dokonaj jednoczynnikowej analizy wariancji ze względu na zmienną utworzoną w pkt. 2 (pamiętaj o wnioskach wyciągniętych w punkcie 4). Nie zapomnij:
  - a) Sprawdzić jednorodność wariancji w grupach (dokonaj oceny wizualnej, wykonaj testy: Flignera-Killeena - funkcja `fligner.test`, oraz Levene'a - funkcja `leveneTest` z pakietu `car`)
  - b) Wykonać wykresy diagnostyczne za pomocą funkcji `plot`.
  - c) Skomentować wyniki otrzymane za pomocą funkcji `summary.aov` i `summary.lm`

```
#A
fligner.test(mouse$weight~mouse$trans,mouse%>%filter(sex=='samica'))

##
## Fligner-Killeen test of homogeneity of variances
##
## data: mouse$weight by mouse$trans
## Fligner-Killeen:med chi-squared = 6.9823, df = 3, p-value = 0.07246
fligner.test(mouse$weight~mouse$trans,mouse%>%filter(sex=='samiec'))

##
```

```
## Fligner-Killeen test of homogeneity of variances
##
## data: mouse$weight by mouse$trans
## Fligner-Killeen:med chi-squared = 6.9823, df = 3, p-value = 0.07246
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##      recode
```

```
leveneTest(mouse$weight~mouse$trans,mouse%>%filter(sex=='samica'))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  3  1.9562 0.1196
##      528
```

```
leveneTest(mouse$weight~mouse$trans,mouse%>%filter(sex=='samiec'))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  3  1.9562 0.1196
##      528
```

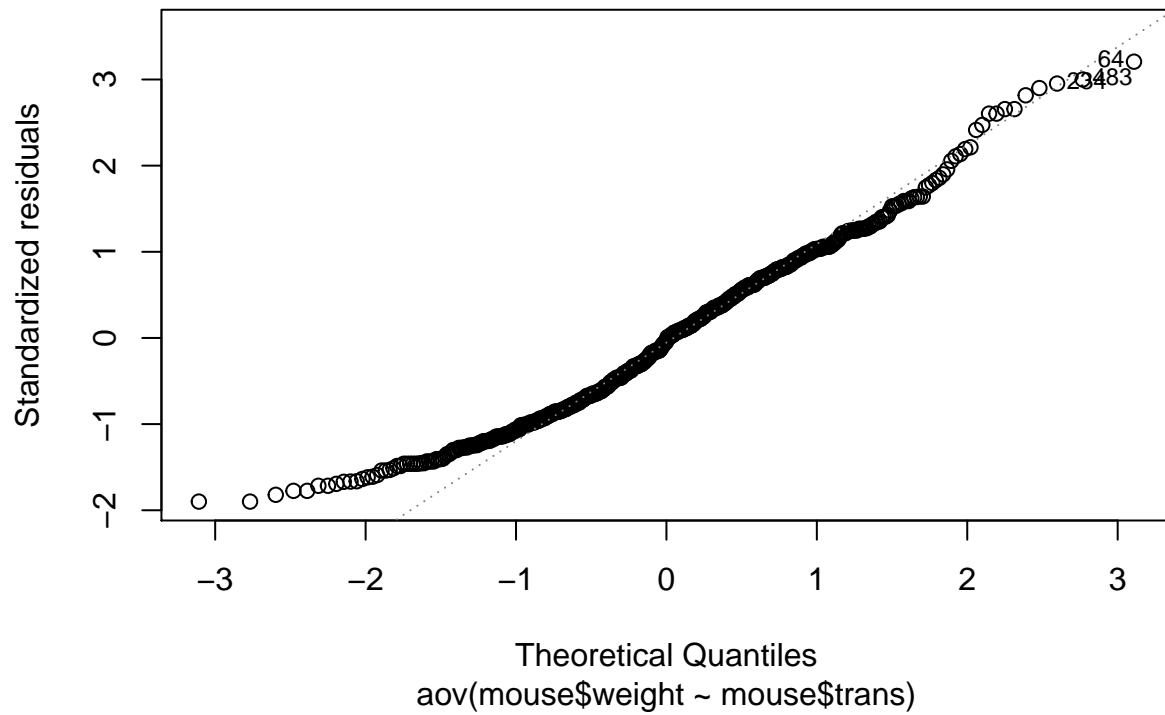
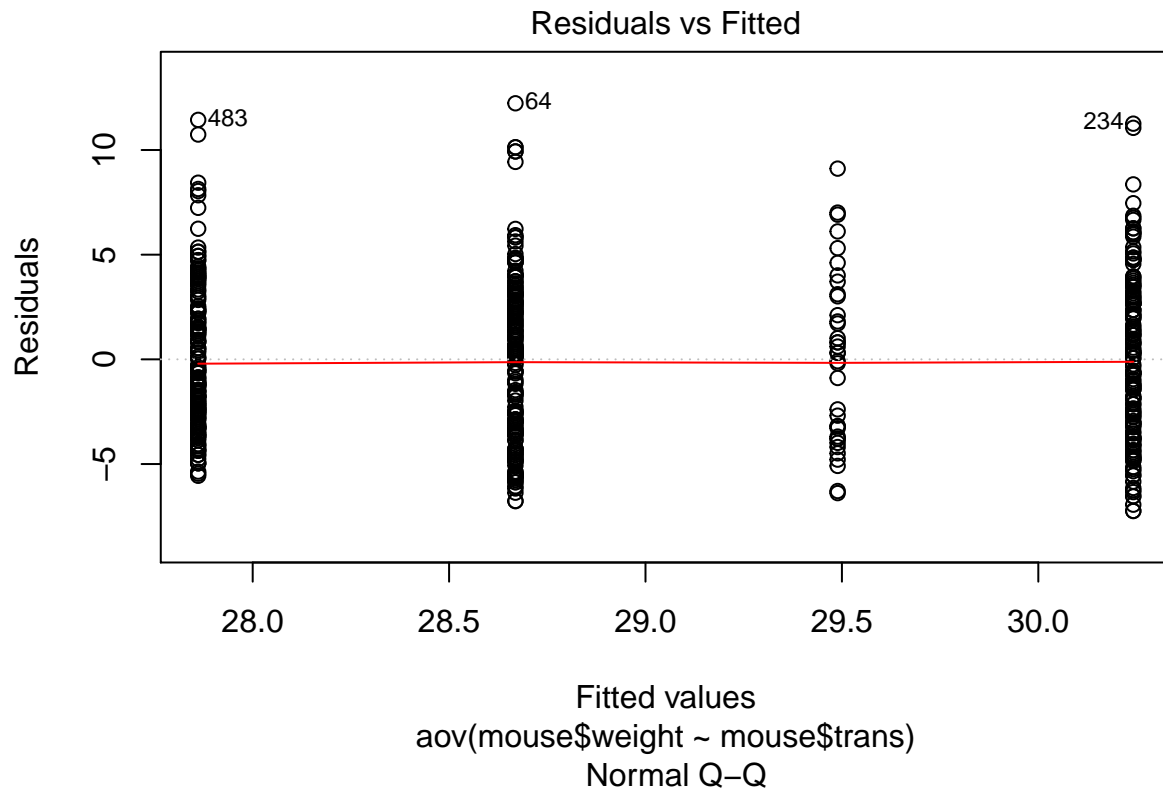
---

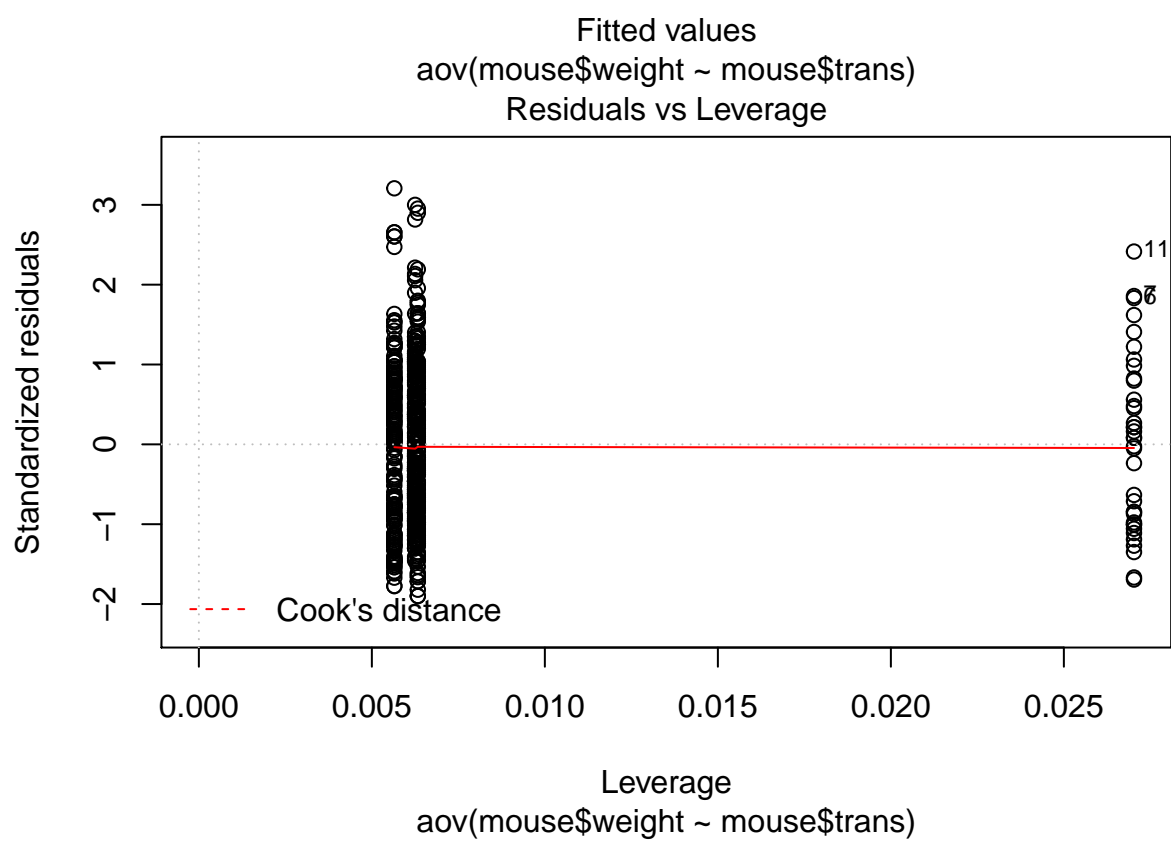
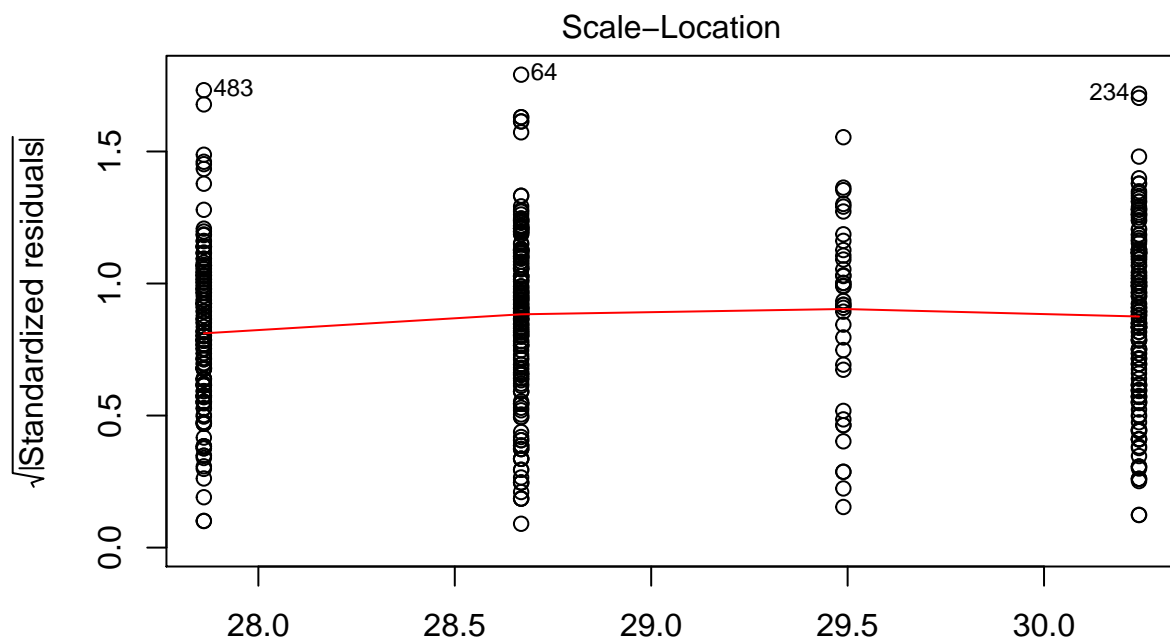
---

```
#B
```

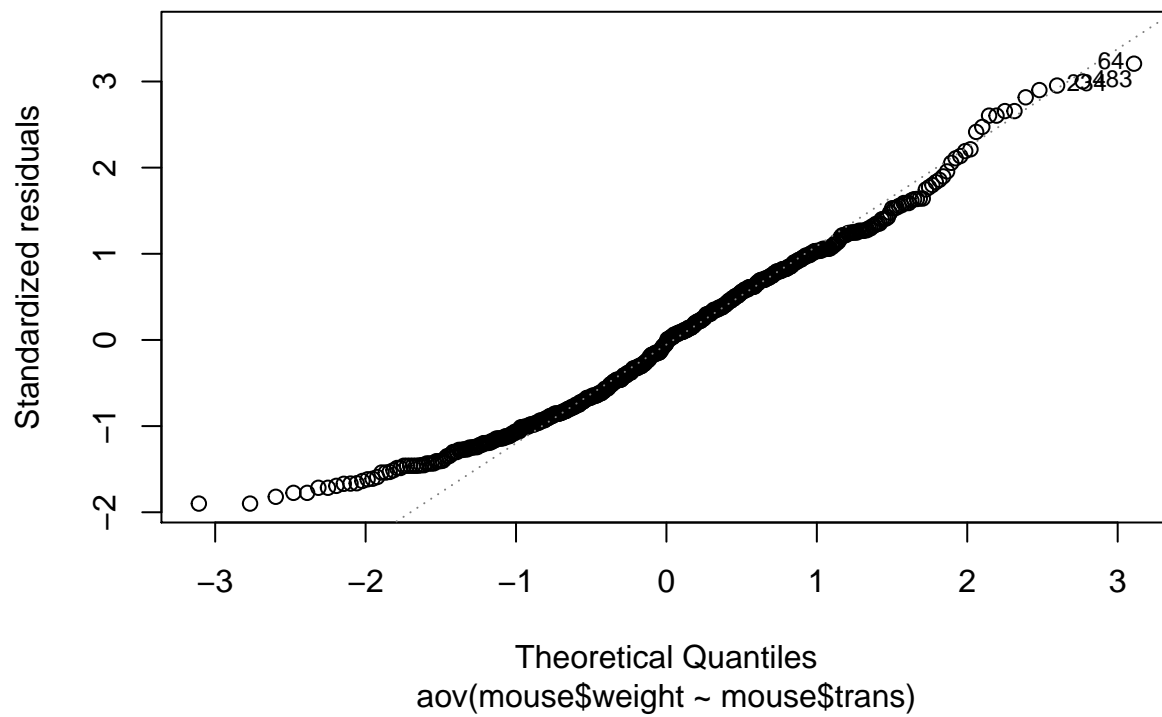
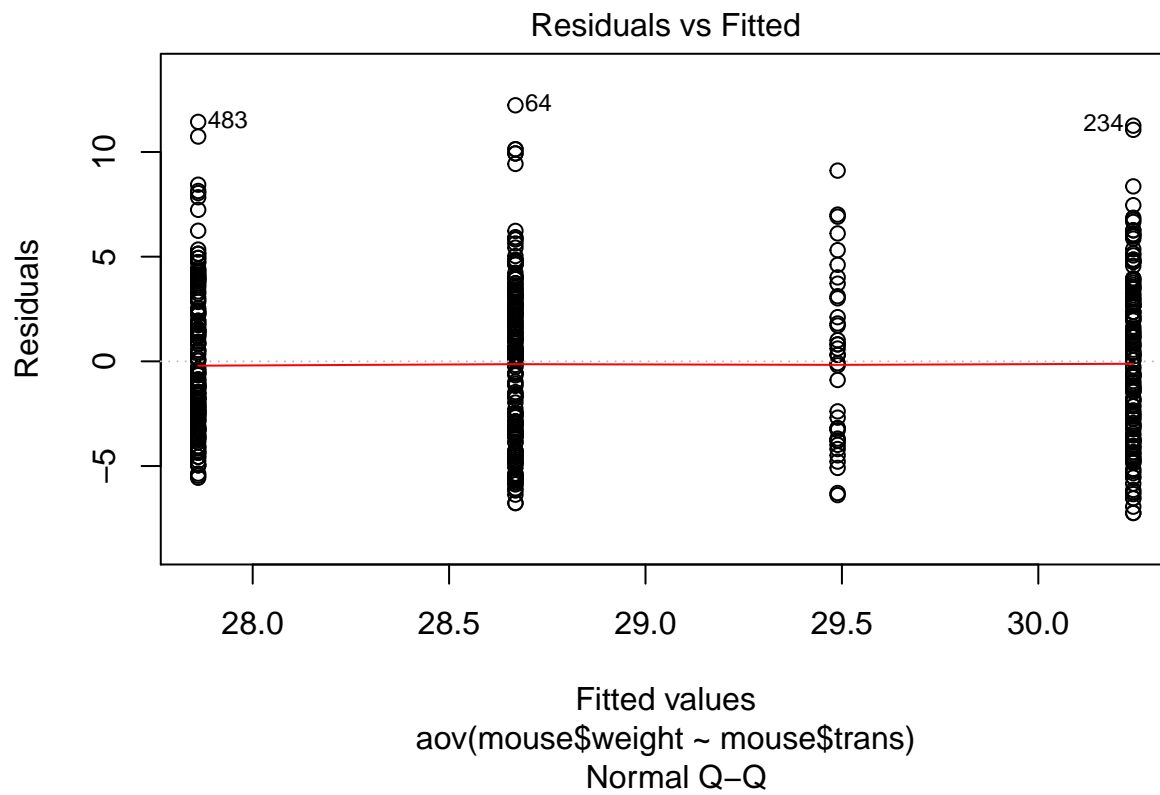
```
aov.samiec=aov(mouse$weight~mouse$trans,mouse%>%filter(sex=='samiec'))
aov.samica=aov(mouse$weight~mouse$trans,mouse%>%filter(sex=='samica'))

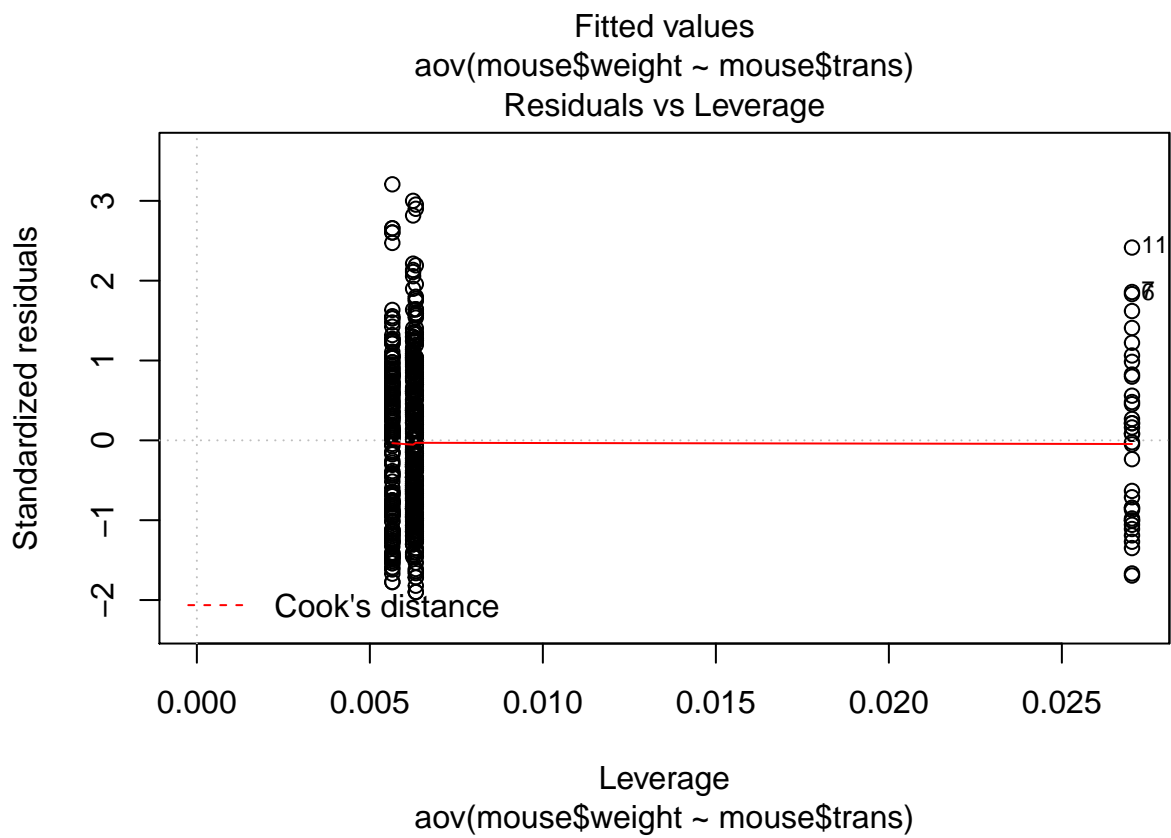
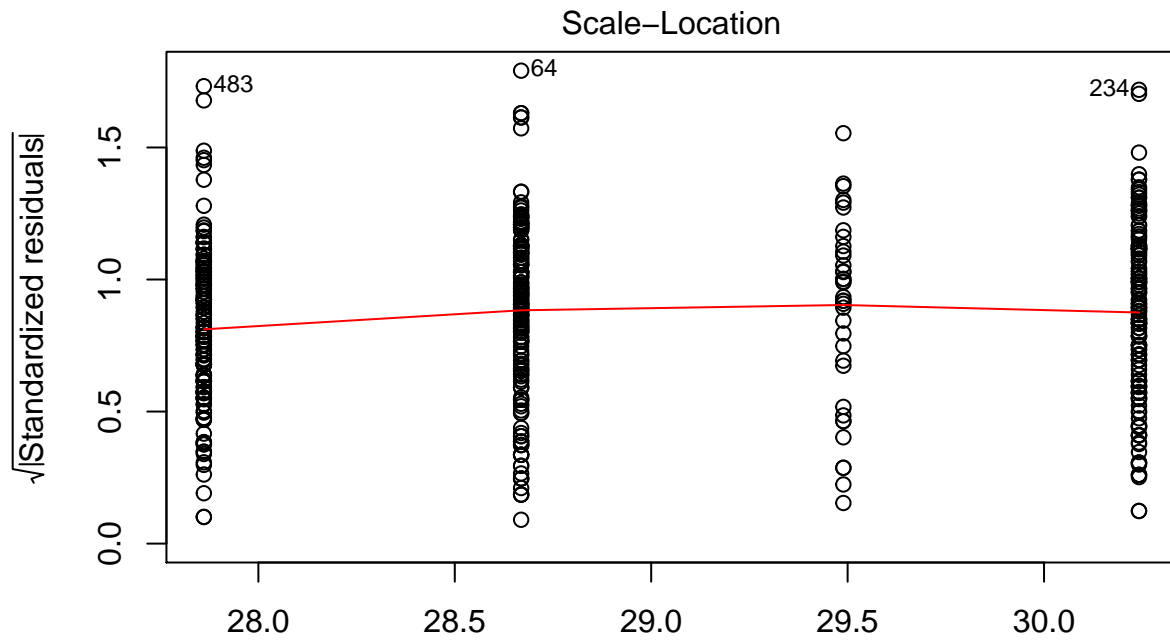
plot(aov.samica)
```





```
plot(aov.samiec)
```





```
#C
```

```
summary(aov.samica)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## mouse$trans  3     478    159.35   10.89 5.95e-07 ***
## Residuals 528    7725     14.63
```



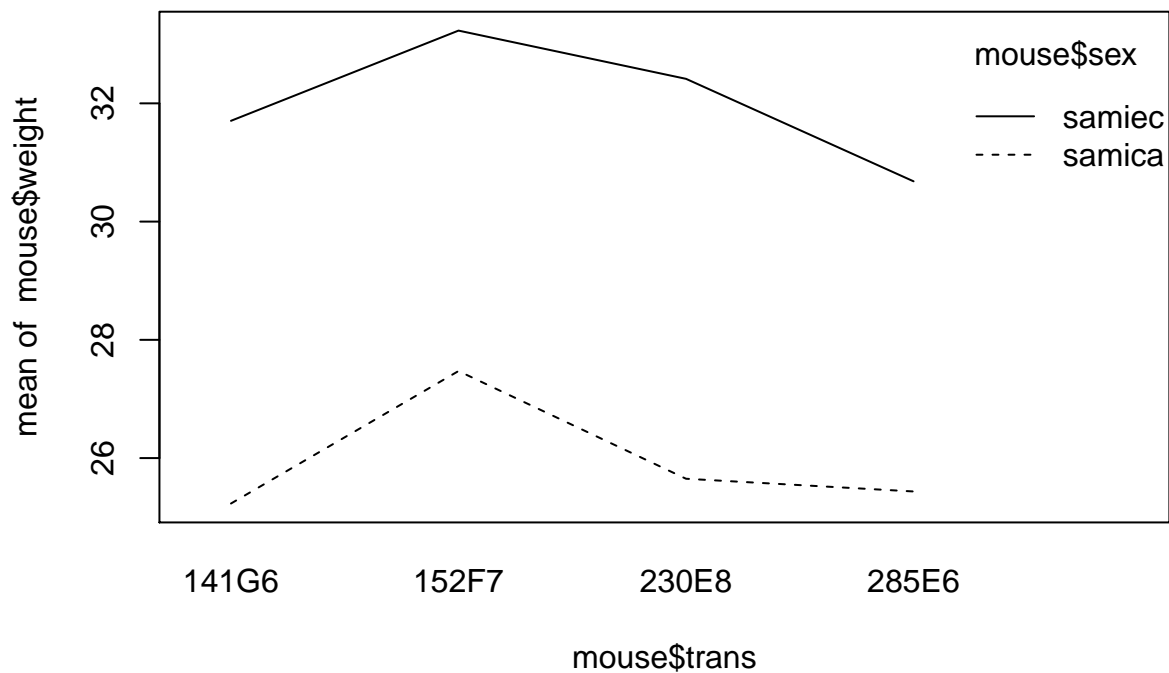
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(aov.samiec)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## mouse$trans   3    478   159.35   10.89 5.95e-07 ***
## Residuals    528    7725    14.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7. Narysuj wykres interakcji (za pomocą `interaction.plot`). Następnie wykonaj dwuczynnikową analizę wariancji (rozważ pełny model z wszystkimi interakcjami oraz model addytywny) ze względu na zmienną z pkt 2 i płeć myszy. Jakie wnioski z niej wypływają?

```
interaction.plot(mouse$trans,mouse$sex,mouse$weight)
```



Myszy z genem 152F7 mają największą wagę.