

Practical 2.2

CHEN

2024-01-23

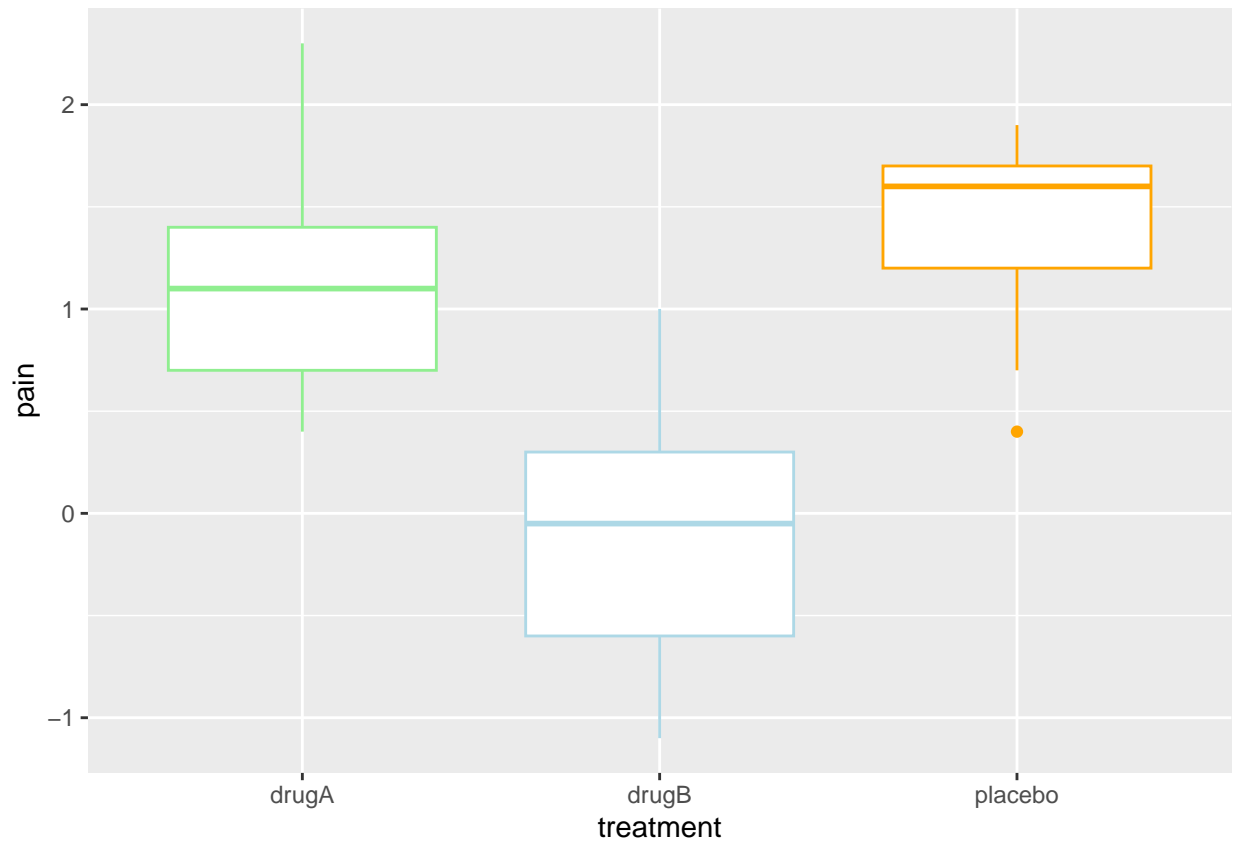
Drug trial dataset - revisited

Formulate H0 and H1

What are the Null and Alternative Hypotheses? Write them down. H0: There is no difference between different drug trial groups in pain score. HA: There is a difference between different drug trial groups in pain score.

Load and inspect the dataset

```
drug <- read.csv("drug_trial.csv")
library(ggplot2)
# Plot the data in a boxplot
g1 <- ggplot(data = drug, aes(x = treatment, y = pain))
g1_1 <- g1 + geom_boxplot(color = c("lightgreen", "lightblue", "orange"))
g1_1
```



Check whether the assumptions for an ANOVA are met

List the three assumptions that need to be met for an ANOVA: 1.Independent random sampling 2.Normality of residuals 3.Equality of variances Do the assessment about whether those assumptions are met.

Independent random sampling

Believe that this is true given the description of the experiment itself.

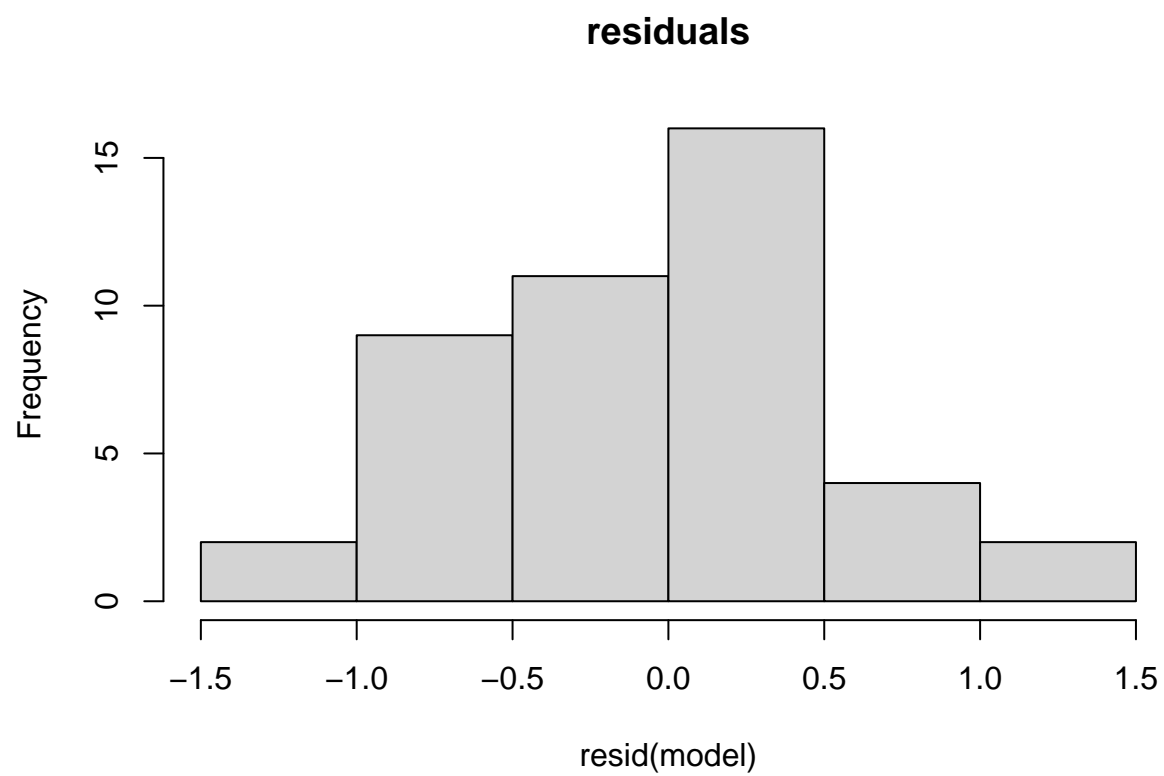
Normality of residuals

Make aov model:

```
model <- aov(pain ~ treatment,
data = drug)
```

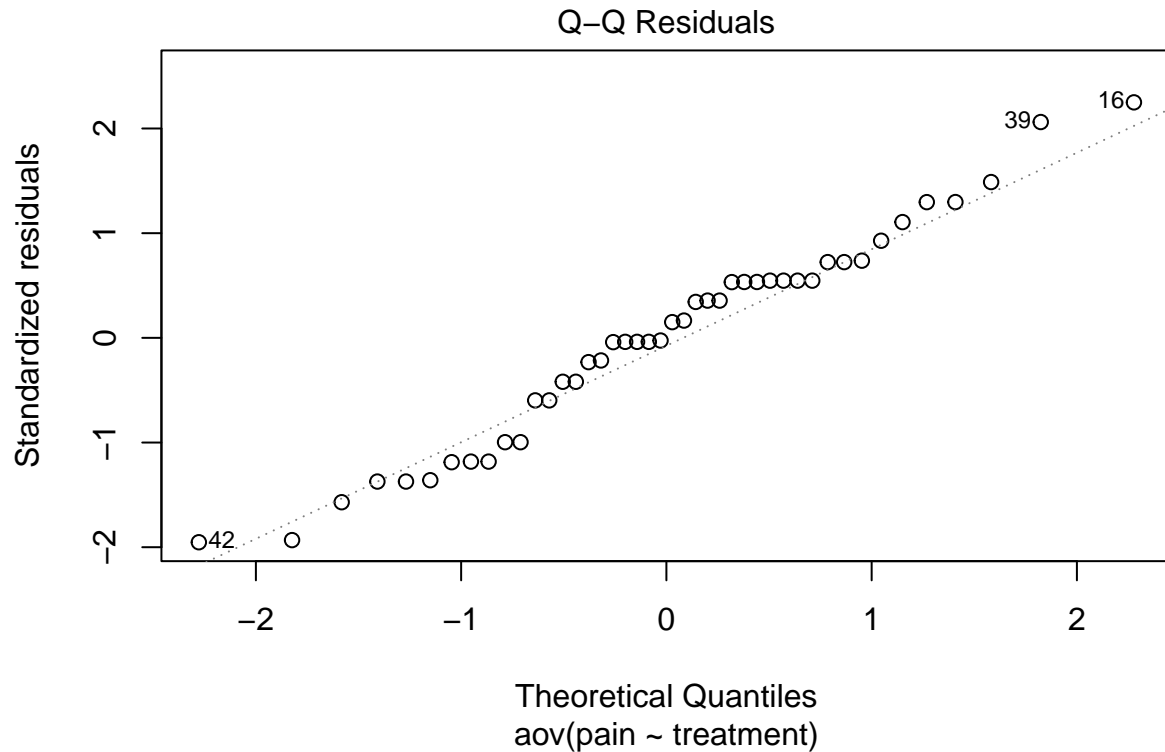
Method 1.

```
hist(resid(model), main = "residuals")
```



Method 2. The plot is the second one shown and is called Normal Q-Q. Dots should be aligned along the diagonal.

```
plot(model, 2)
```



Method 3. Use a formal test for normality, e.g. the Shapiro-Wilk test H_0 : H_A :

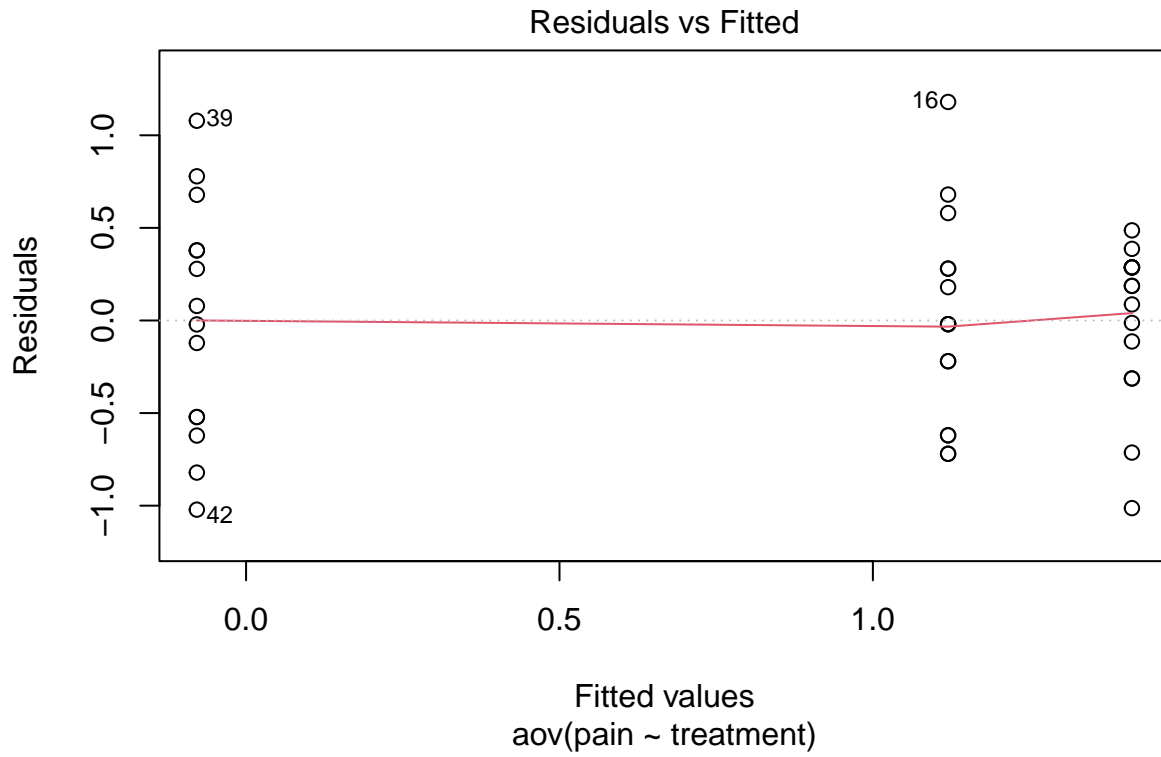
```
shapiro.test(resid(model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model)
## W = 0.97441, p-value = 0.4286
```

Equality of variances

Use “Residuals vs Fitted” plot. Looking for similar heights of “columns”.

```
plot(model, 1)
```



Run an ANOVA

```
summary(model)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment   2  17.92   8.960    30.4 8.01e-09 ***
## Residuals  41   12.09   0.295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
gg <- ggplot(data=mouse, mapping=aes(x=genotype, y=weight_gain))
gg1 <- gg+geom_boxplot()
gg1
```

Run a post-hoc test

It looks like we have a significant result, so now it is time for a post-hoc test.

```
TukeyHSD(model)
```

```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = pain ~ treatment, data = drug)
```

```
##
## $treatment
##           diff           lwr           upr           p adj
## drugB-drugA -1.1985714 -1.6891632 -0.7079797 0.0000016
## placebo-drugA 0.2933333 -0.1887258 0.7753924 0.3110505
## placebo-drugB 1.4919048 1.0013130 1.9824965 0.0000000
```

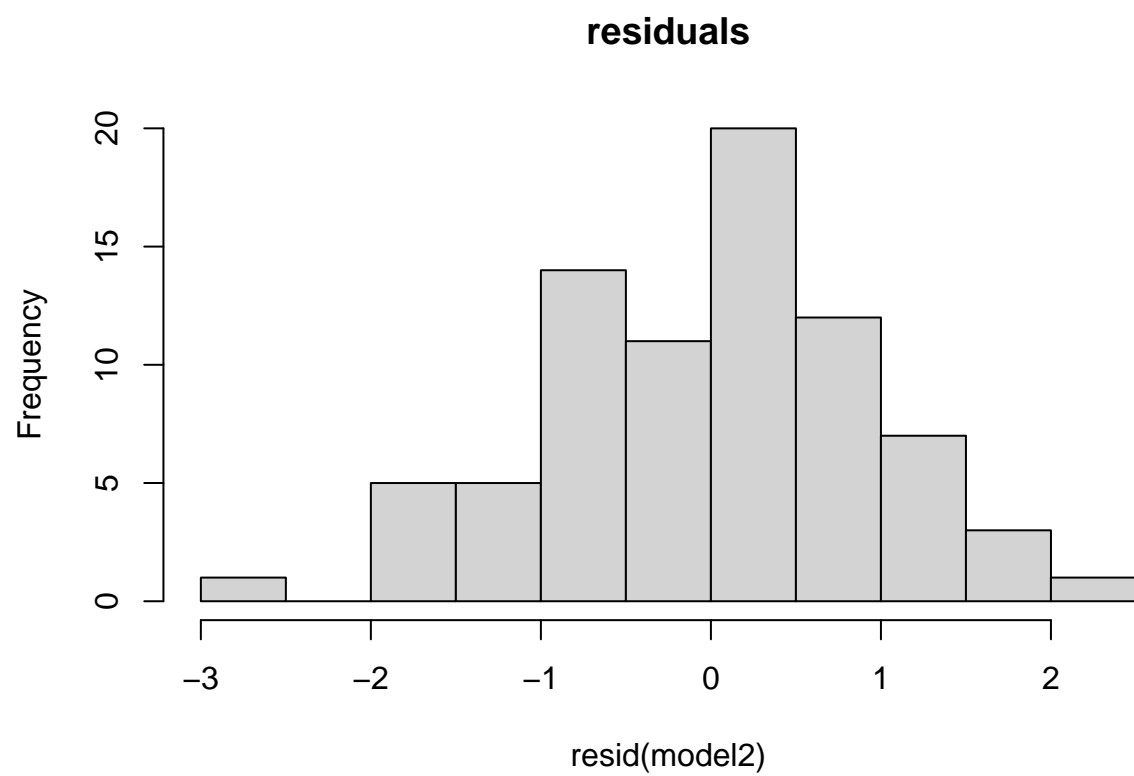
The “p adj” values are the adjusted p values. This means that they have already been corrected for multiple testing, so you can go ahead and compare those values directly to the alpha level of your choice. What do you conclude?

Weight loss in mice

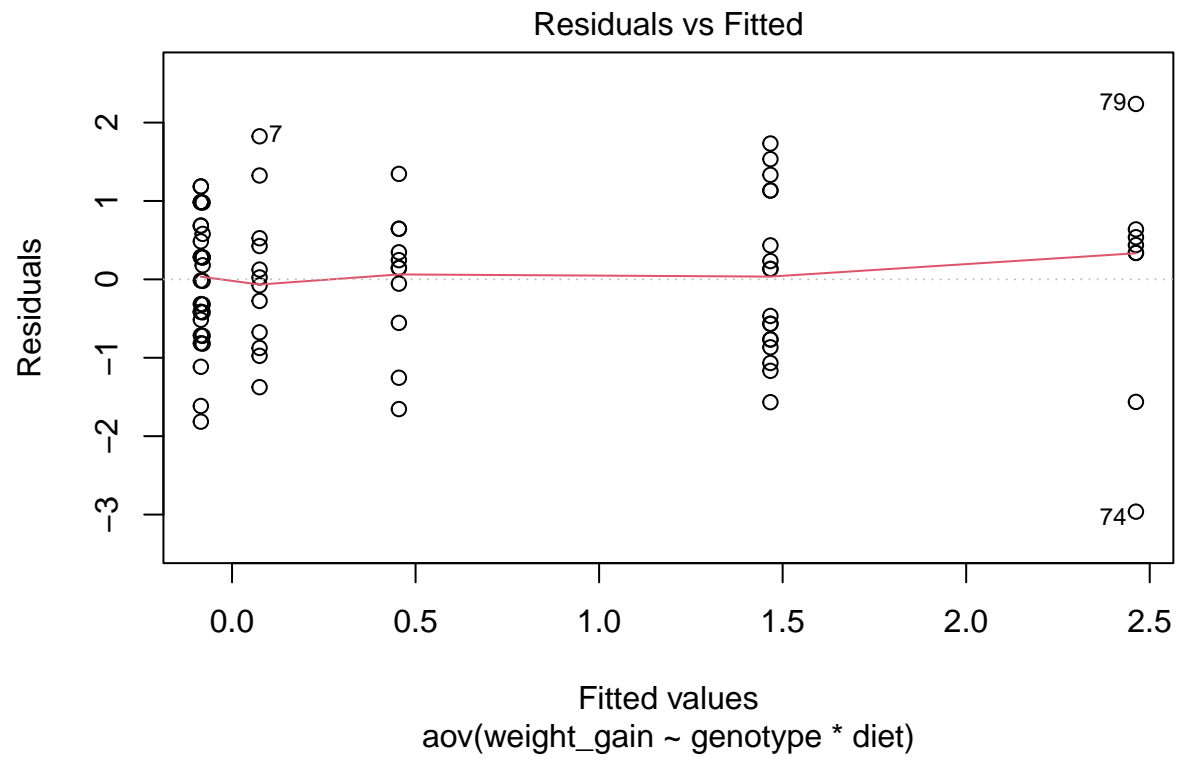
You are interested in a gene that is said to control appetite control. Specifically, mice with one or two copies of the “A” allele seem to be better than mice with the “B” allele at controlling their food intake, even if unlimited food is available. Is this appetite control allele also related to weight loss? For this experiment, you tested mice of the homozygous “AA” and “BB” genotypes, as well as heterozygous “AB” mice. Mice were randomly separated into two groups: One group had access to only as much food as would be required to maintain their weight (“restricted”). The other group has access to unlimited food (“unrestricted”). Mice were weighed at the beginning and the end of a 1-month experiment, and their weight gain (in g) recorded (this number can be negative if the animal has lost weight). The data are provided in file `mouse_experiment.csv`. Import, visualise, and analyse the dataset.

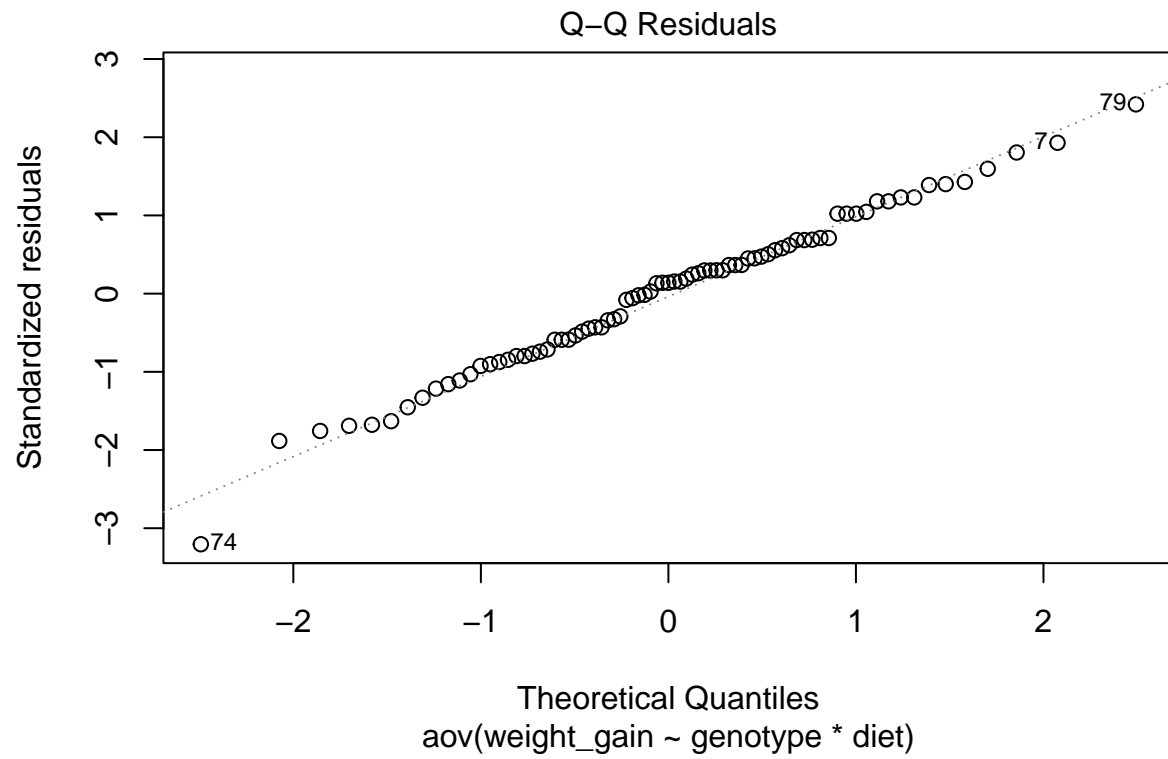
```
mouse <- read.csv("mouse_experiment.csv")

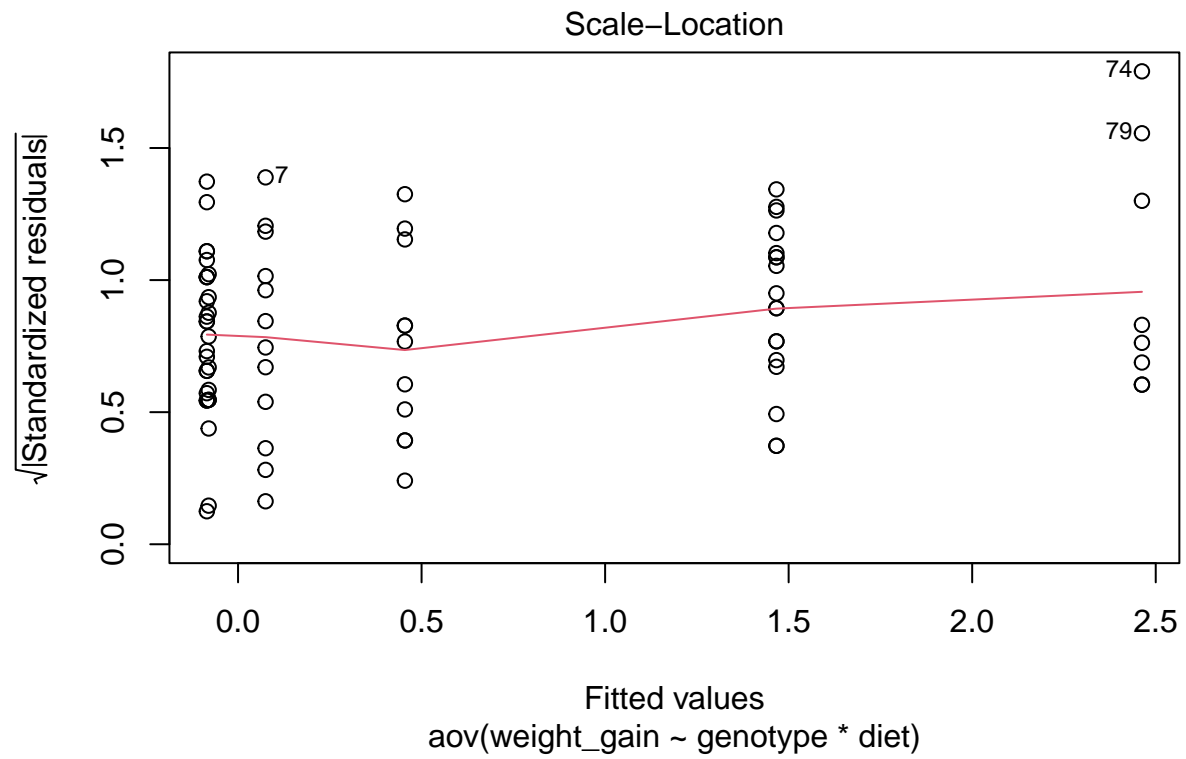
model2 <- aov(weight_gain ~ genotype * diet,
data = mouse)
hist(resid(model2),main = "residuals")
```

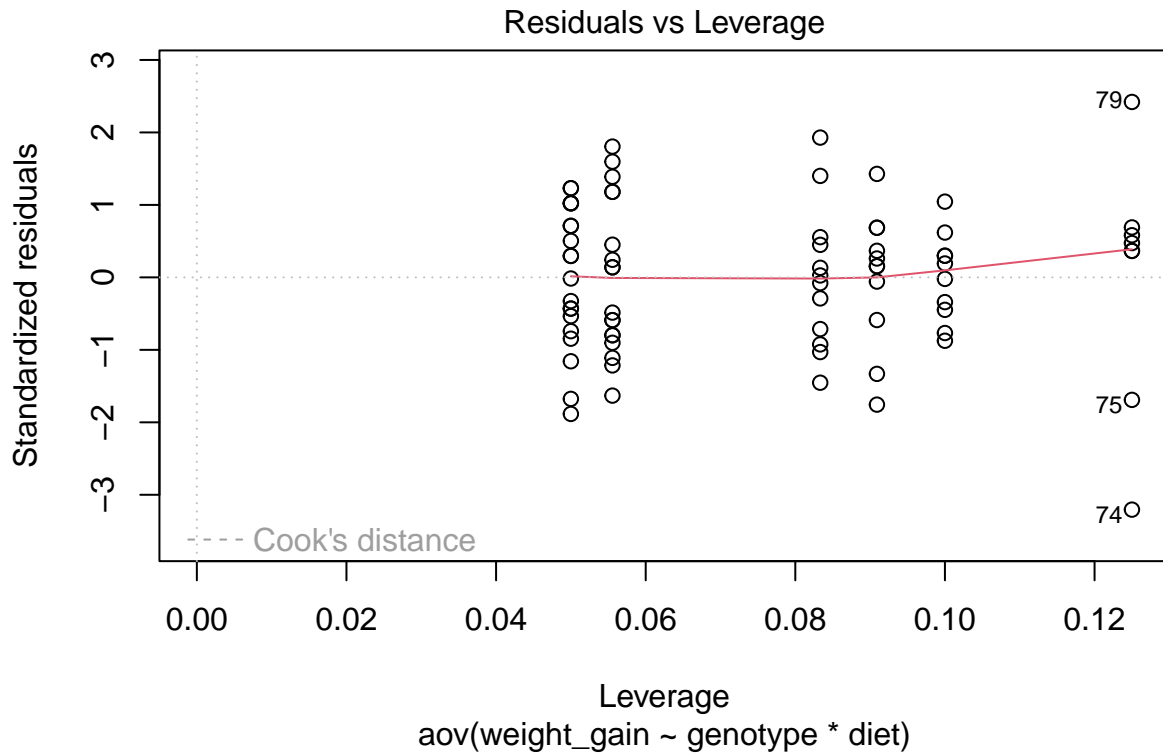


```
plot(model2)
```









```
summary(model2)
##              Df Sum Sq Mean Sq F value    Pr(>F)
## genotype      2   6.40     3.20   3.276 0.04342 *
## diet          1  40.39    40.39  41.362 1.16e-08 ***
## genotype:diet  2  11.97     5.99   6.130 0.00346 **
## Residuals     73  71.29     0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
TukeyHSD(model2)
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = weight_gain ~ genotype * diet, data = mouse)
##
## $genotype
##           diff           lwr           upr           p adj
## AB-AA 0.3934783 -0.2311222  1.018079  0.2935392
## BB-AA 0.7934783  0.0494578  1.537499  0.0338875
## BB-AB 0.4000000 -0.2764864  1.076486  0.3387304
##
## $diet
##           diff           lwr           upr p adj
## unrestricted-restricted 1.432486 0.9884195 1.876552 0
##
## $`genotype:diet`
##           diff           lwr           upr           p adj
```

## AB:restricted-AA:restricted	-0.1600000	-1.21611661	0.8961166	0.9977516
## BB:restricted-AA:restricted	-0.1550000	-1.39340649	1.0834065	0.9991025
## AA:unrestricted-AA:restricted	0.3795455	-0.82776637	1.5868573	0.9401208
## AB:unrestricted-AA:restricted	1.3916667	0.31377217	2.4695612	0.0041690
## BB:unrestricted-AA:restricted	2.3875000	1.06735424	3.7076458	0.0000176
## BB:restricted-AB:restricted	0.0050000	-1.11518082	1.1251808	1.0000000
## AA:unrestricted-AB:restricted	0.5395455	-0.54615997	1.6252509	0.6937938
## AB:unrestricted-AB:restricted	1.5516667	0.61198003	2.4913533	0.0001032
## BB:unrestricted-AB:restricted	2.5475000	1.33756643	3.7574336	0.0000005
## AA:unrestricted-BB:restricted	0.5345455	-0.72918895	1.7982799	0.8167215
## AB:unrestricted-BB:restricted	1.5466667	0.40593036	2.6874030	0.0022395
## BB:unrestricted-BB:restricted	2.5425000	1.17056429	3.9144357	0.0000105
## AB:unrestricted-AA:unrestricted	1.0121212	-0.09478011	2.1190225	0.0925815
## BB:unrestricted-AA:unrestricted	2.0079545	0.66402047	3.3518886	0.0005552
## BB:unrestricted-AB:unrestricted	0.9958333	-0.23315548	2.2248221	0.1800034