Introduction to R: basic programming

Ziv Shkedy, Hasselt University

```
## Warning: package 'mvtnorm' was built under R version 3.6.2
```

Introduction

Slides, code and tutorials

This part of the interactive book contains all R code that was used to produce the results and output presented in chapter 3 (modeling:one way ANOVA) in the course's slides. We include YouTube tutorials as a part of the book and links to the relevant tutorials are provided in different chapters. Note that these tutorials were not developed especially for this book, they cover the same topics using different examples.

R ?

No previous knowledge about R is required. We use the R function aov() to fit a One-Way ANOVA in R and the chicken Weights dataset is used for illustraion. The model can be fitted using the functions lm() and glm() as well.

Slides

Slide for this part of the course are avilable online in the >eR-BioStat website. See Rcoursemodeling.

The Chicken Weights data

Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement. Their weights (the response variable) in grams after six weeks are given along with feed types (the factor). The Chicken Weights data is a data frame in R, chickwts .

```
head(chickwts)
```

```
##   weight      feed
## 1    179 horsebean
## 2    160 horsebean
## 3    136 horsebean
## 4    227 horsebean
## 5    217 horsebean
## 6    168 horsebean
```

One-Way ANOVA model

YouTube tutorials: One-Way ANOVA in R

R - One-way ANOVA

For a YouTube tutorial about One-Way ANOVA in R R Statistics and Research see YTOneWayANOVA1.

One way ANOVA in RStudio

For a YouTube tutorial about One-Way ANOVA by Tom Sherrattin R see YTOneWayANOVA2.

Model formulation

We consider a one-way ANOVA model for that data

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1, \ldots, 6,\ j = 1, \ldots, n_i.$$

Here,

- Y_{ij} is the weight of the j'th subject in diet group i.
- The parameters μ_i represent the mean of the distribution of weight at each age group.
- \item ε_{ij} is a random error which assumed to be normally distributed,

  $$\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$

  .

Our primary of interest is to estimate the group means and then to test the hypotheses

$$\begin{aligned} &\mathrm{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_6, \\ &\mathrm{H}_1 : \mu_l \neq \mu_k \ \text{ for at least one pair.} \end{aligned}$$

Visualizing the Data

The striptplot in Figure~@ref(fig:fig1) reveals that the cash offers in the middle age group are higher than the cash offers in the young and elderly age groups.

```
stripplot(feed ~ jitter(weight),
            data = chickwts,
            aspect = 1, jitter = T,
            xlab="Cash Offers", col = 1)
```

The boxplot in Figure~@ref(fig:fig2) can be used to visualize the patterns in teh data. Note how the box of the middle group is located higher than the boxes of the other groups.

```
boxplot(split(chickwts$weight,chickwts$feed))
```

The sample means for the 6 diet groups are equal to

```
 tapply(chickwts$weight,chickwts$feed,mean)
```

```
##    casein horsebean   linseed  meatmeal   soybean sunflower
##  323.5833  160.2000  218.7500  276.9091  246.4286  328.9167
```

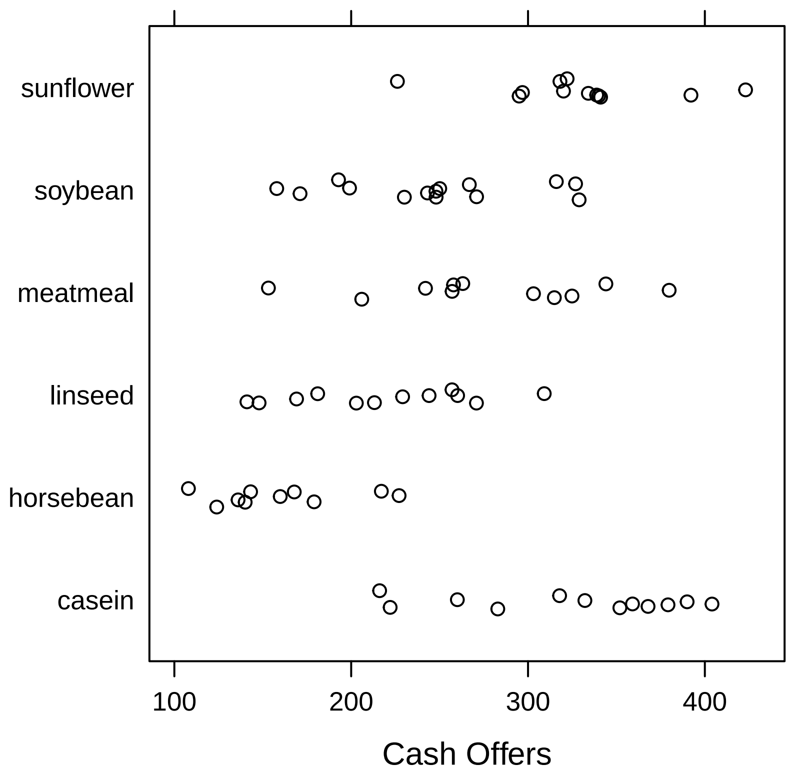


Figure 1: Chicks weight by diet group.

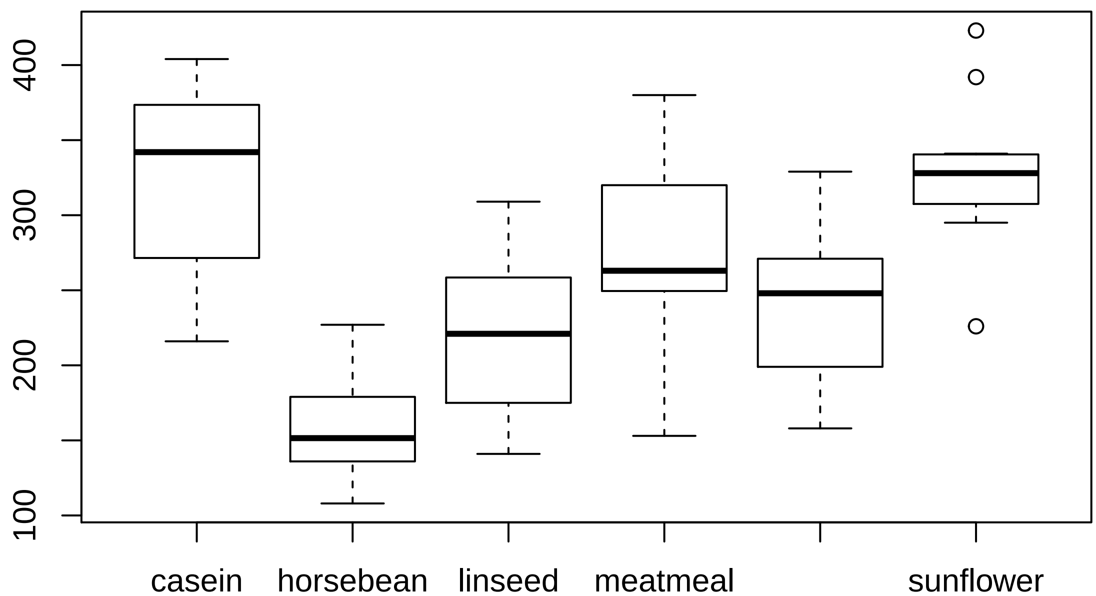


Figure 2: Chicks weight by diet group.

Fitting a One-way ANOVA Model in R

The aov() Function

The R function which we use to fit an ANOVA model in R is aov() . A General call of the function has the form of aov(dependent variable~factor) . For example one-way ANOVA for the response y and the factor x can be fitted using aov(y~x) . If x is a numerical vector, we can use aov(y~as.factor(x)) .

Fitting an ANOVA Modelfor the Chicken Weights data

In order to fit the model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \;\; i = 1, \ldots, n,$$

we use

```
Fit.aov<-aov(chickwts$weight~chickwts$feed)
```

the object Fit.aov contains the results. The ANOVA table is given by:

```
summary(Fit.aov)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)    
## chickwts$feed  5 231129   46226   15.37 5.94e-10 ***
## Residuals     65 195556    3009                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostic Plots

Residuals

The object Fit.aov$resid contains the residuals values.

```
Fit.aov$resid
```

```
##           1           2           3           4           5           6 
##   18.800000   -0.200000  -24.200000   66.800000   56.800000    7.800000 
##           7           8           9          10          11          12 
##  -52.200000  -36.200000  -17.200000  -20.200000   90.250000   10.250000 
##          13          14          15          16          17          18 
##  -37.750000  -77.750000   41.250000  -15.750000  -70.750000  -49.750000 
##          19          20          21          22          23          24 
##   -5.750000   38.250000   25.250000   52.250000   -3.428571  -16.428571 
##          25          26          27          28          29          30 
##    1.571429   80.571429   82.571429    3.571429  -53.428571   24.571429 
##          31          32          33          34          35          36 
##   69.571429   20.571429  -47.428571  -75.428571  -88.428571    1.571429 
##          37          38          39          40          41          42 
##   94.083333   11.083333   63.083333   10.083333   12.083333 -102.916667 
##          43          44          45          46          47          48 
##   -8.916667  -33.916667    5.083333   -6.916667  -31.916667  -10.916667 
##          49          50          51          52          53          54 
##   48.090909  -19.909091   26.090909   38.090909  103.090909 -123.909091 
```

```
##          55          56          57          58          59          60
##  -13.909091  -34.909091  -70.909091   67.090909  -18.909091   44.416667
##          61          62          63          64          65          66
##   66.416667   55.416667  -63.583333   80.416667   -5.583333   28.416667
##          67          68          69          70          71
##   35.416667 -107.583333 -101.583333  -40.583333    8.416667
```

Stem-and-leaf diagram reveals a symatical distribution.

```r
stem(Fit.aov$resid)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   -12 | 4
##   -10 | 832
##    -8 | 8
##    -6 | 85114
##    -4 | 32071
##    -2 | 86542400
##    -0 | 976641976630
##     0 | 22458800129
##     2 | 15568588
##     4 | 148257
##     6 | 36770
##     8 | 01304
##    10 | 3
```

Stripplot for the residuals by diet group is shown in Figure~@ref(fig:fig3)

```r
stripplot(chickwts$feed ~ jitter(Fit.aov$resid), aspect = 1, jitter = T, xlab =
    "Residuals", col = 1)
```

Boxplot, histogram and normal probability plot and shown in Figure ~@ref(fig:fig4).

```r
par(mfrow = c(2, 2))
hist(Fit.aov$resid)
boxplot(split(Fit.aov$resid, chickwts$feed))
qqnorm(Fit.aov$resid)
```

Alternativly, these figures can be produce using the function plot() in the following way

```r
plot(Fit.aov)
```

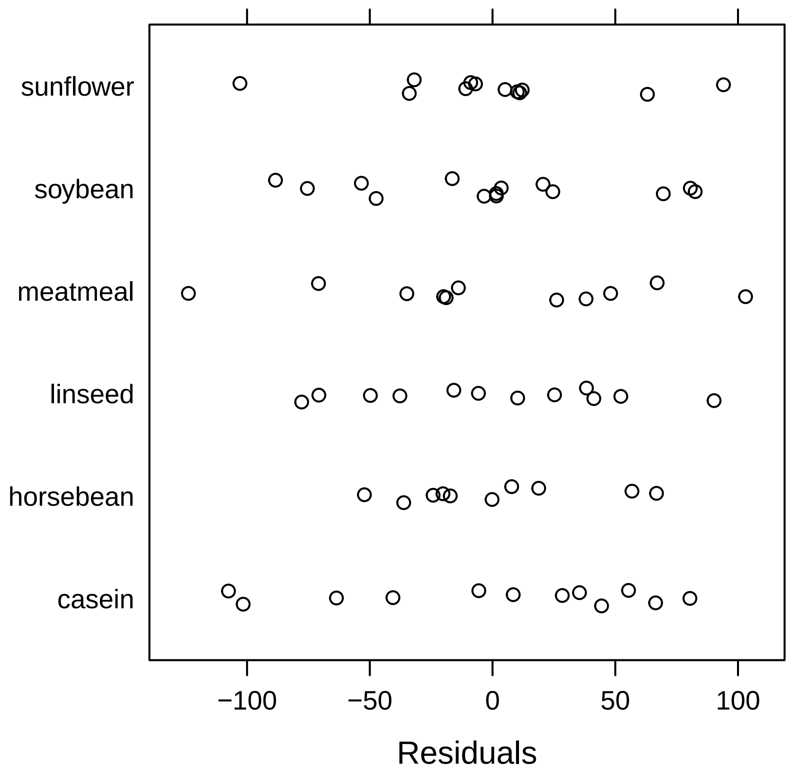


Figure 3: Residuals by diet group.

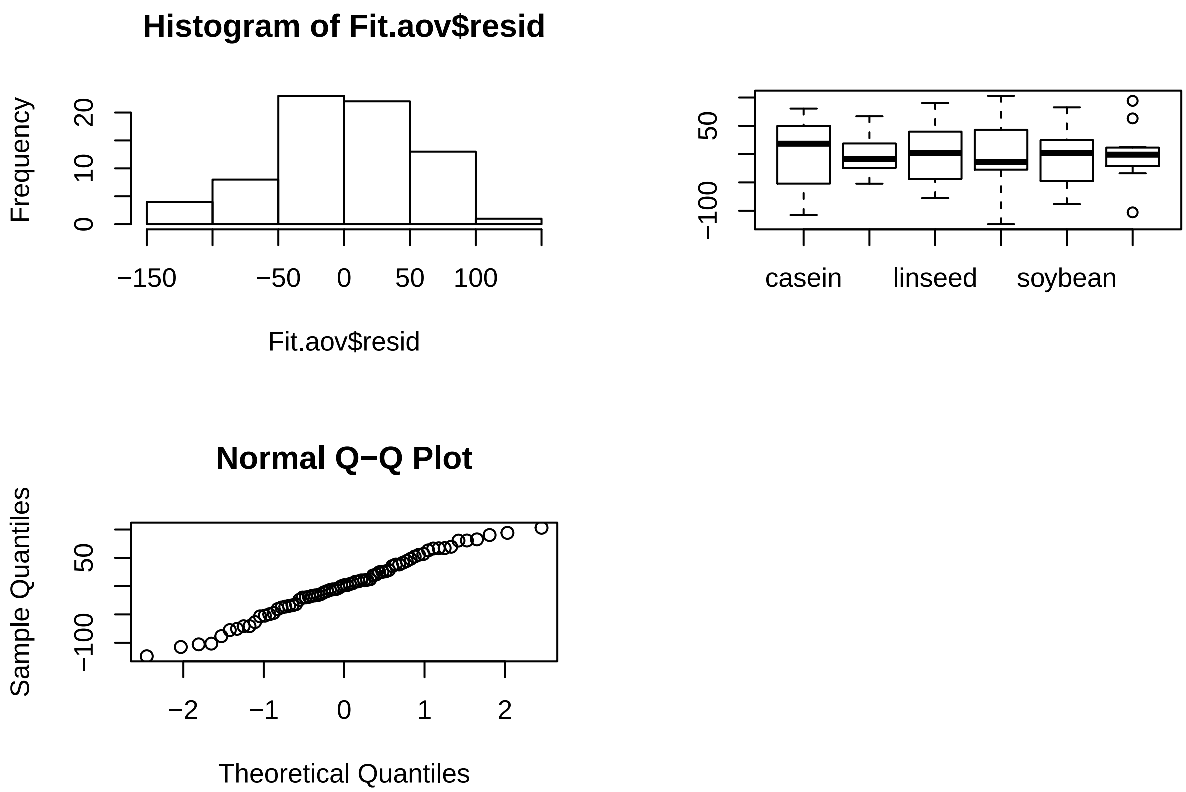
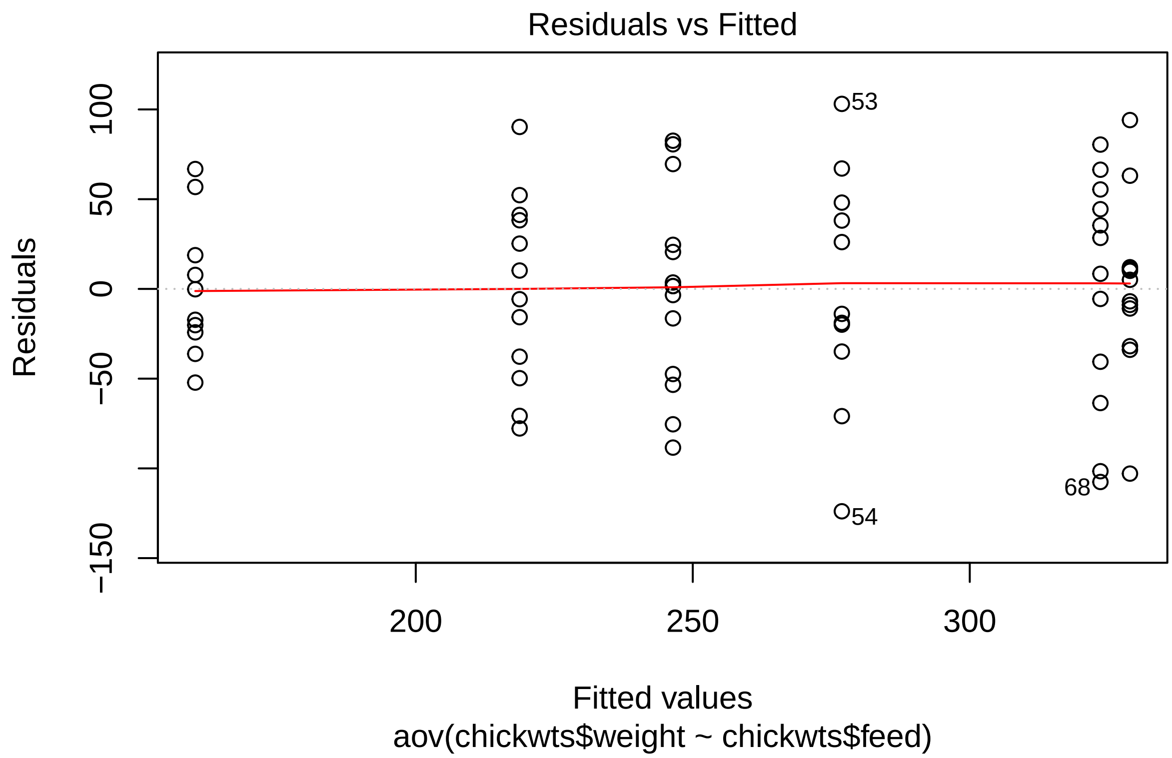
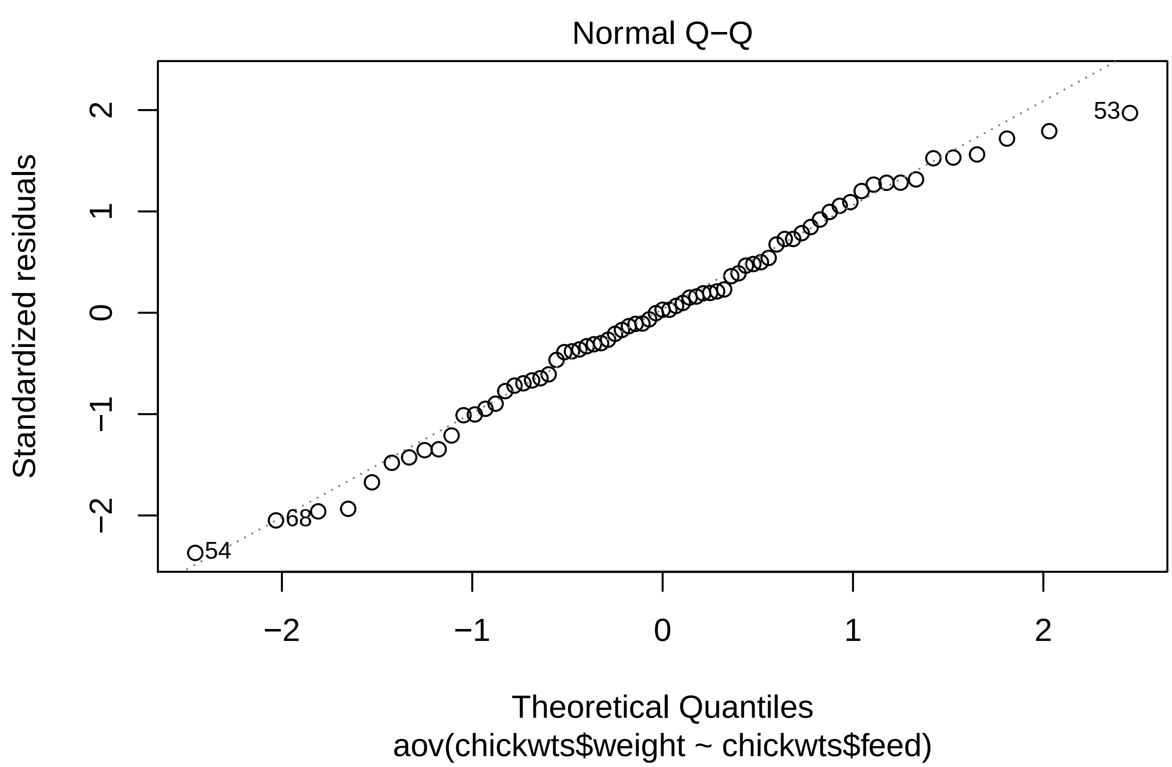
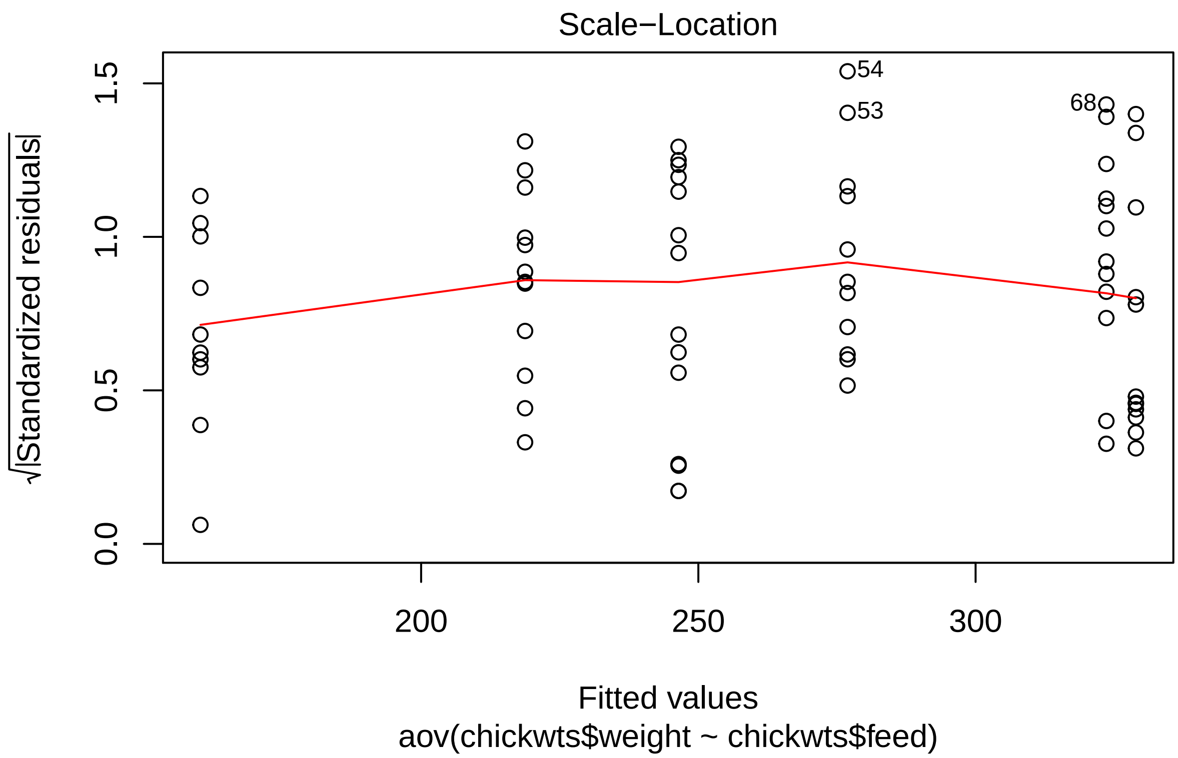
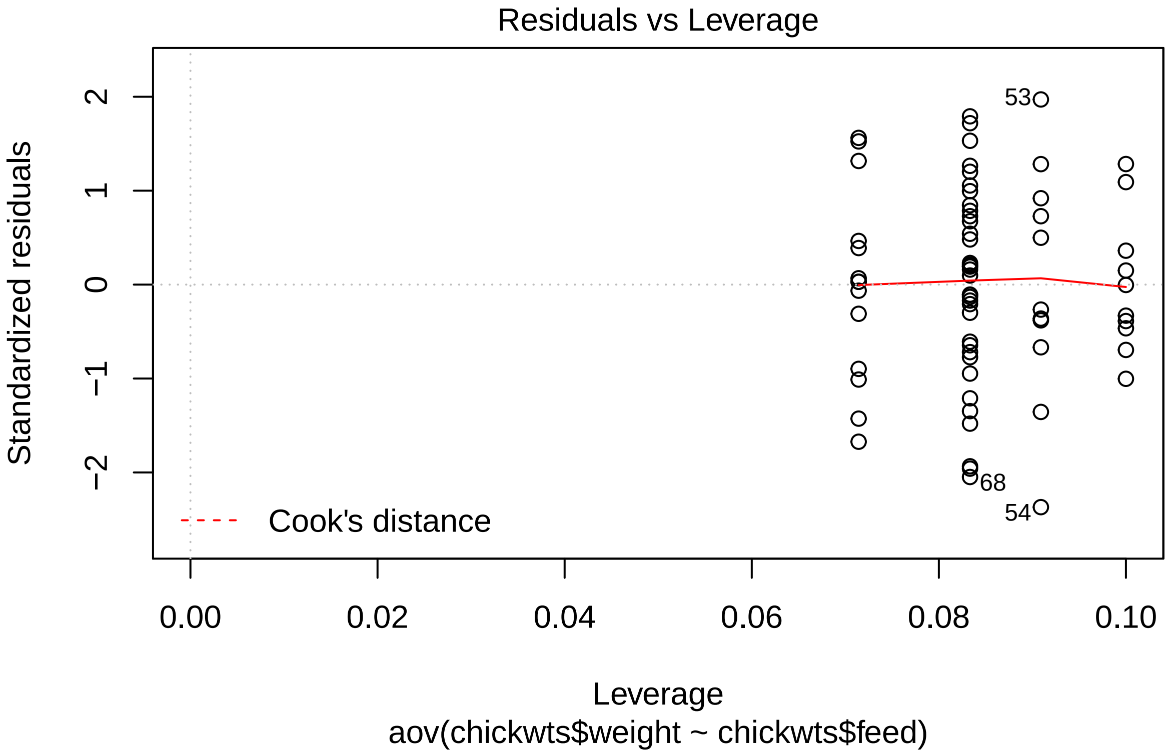


Figure 4: Distribution for the residuals.


Residuals vs Fitted
Residuals
Fitted values
aov(chickwts$weight ~ chickwts$feed)
100
50
0
−50
−150
200
250
300
53
54
68


Normal Q−Q
Standardized residuals
Theoretical Quantiles
aov(chickwts$weight ~ chickwts$feed)
2
1
0
−1
−2
53
68
54


Scale−Location
54
53
68
1.5
1.0
0.5
0.0
√|Standardized residuals|
200
250
300
Fitted values
aov(chickwts$weight ~ chickwts$feed)


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
53
68
54
Cook's distance
0.00
0.02
0.04
0.06
0.08
0.10
Leverage
aov(chickwts$weight ~ chickwts$feed)