

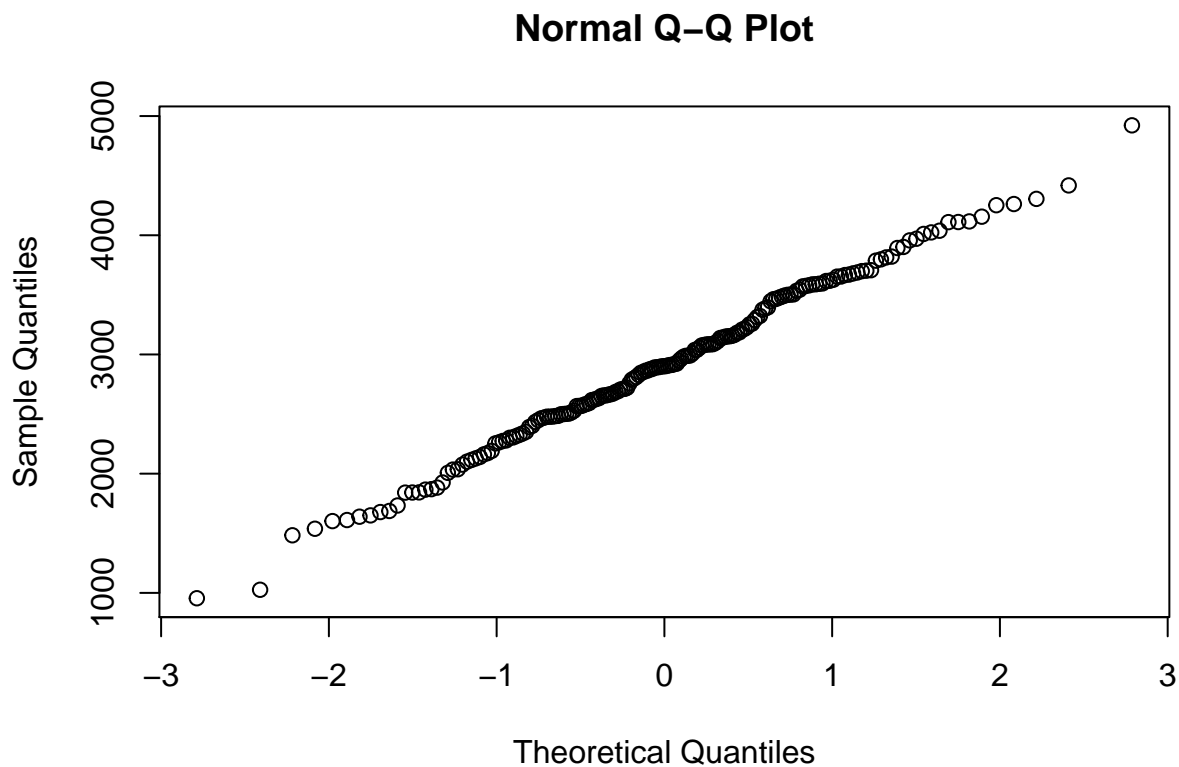
# Assignment 01

EDDA-Group43

2021/2/19

## Exercise 1. Birthweights

```
data = read.table(file="birthweight.txt", header=TRUE)
qqnorm(data$birthweight)
```



```
data_mean = mean(data$birthweight)
error = qt(0.95, df=length(data$birthweight)) * sd(data$birthweight) / sqrt(length(data$birthweight))
upper_bound = data_mean + error
lower_bound = data_mean - error
```

```
t.test(data$birthweight, conf.level=0.9)
```

```
##
## One Sample t-test
```

```
##
## data: data$birthweight
## t = 57.269, df = 187, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
## 2829.202 2997.384
## sample estimates:
## mean of x
## 2913.293

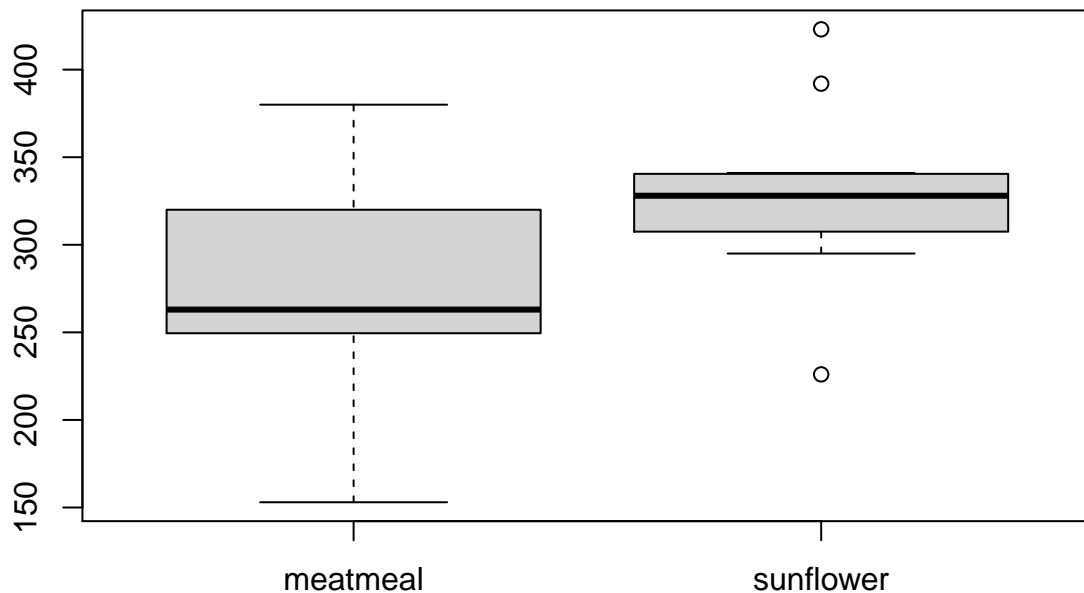
t.test(data$birthweight, mu=2800, alternative="greater")

##
## One Sample t-test
##
## data: data$birthweight
## t = 2.2271, df = 187, p-value = 0.01357
## alternative hypothesis: true mean is greater than 2800
## 95 percent confidence interval:
## 2829.202 Inf
## sample estimates:
## mean of x
## 2913.293
```

## Exercise 5. Chick weights

a.

```
data("chickwts")
meatmeal = chickwts$weight[chickwts$feed == "meatmeal"]
sunflower = chickwts$weight[chickwts$feed == "sunflower"]
boxplot(meatmeal, sunflower, names=c("meatmeal", "sunflower"))
```



```
t.test(meatmeal, sunflower, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  meatmeal and sunflower
## t = -2.1564, df = 18.535, p-value = 0.04441
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -102.572435  -1.442716
## sample estimates:
## mean of x mean of y
##  276.9091  328.9167
```

Since the Welch Two Sample t-test has  $p\text{-value}=0.04441 < 0.05$ , it is the case to reject  $H_0$ . The conclusion should be there exist certain difference between the meatmeal and sunflower. Additionally, the data are not paired, because if we set “paired=TRUE”, we would get an error message- “not all arguments have the same length”.

```
wilcox.test(meatmeal, sunflower)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  meatmeal and sunflower
## W = 36, p-value = 0.06882
## alternative hypothesis: true location shift is not equal to 0
```

Mann-Whitney test has a  $p$ -value = 0.06882 > 0.05, so the conclusion is that there is no such significant difference between two groups.

```
ks.test(meatmeal,sunflower)
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: meatmeal and sunflower
## D = 0.47727, p-value = 0.1085
## alternative hypothesis: two-sided
```

Kolmogorov-Smirnov test has a  $p$ -value = 0.1085 < 0.05. From this result we can conclude that the two populations are extremely unsymmetrical in shapes.

b.

```
chickaov=lm(weight ~ feed, data=chickwts)
anova(chickaov)
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## feed         5 231129   46226  15.365 5.936e-10 ***
## Residuals   65 195556     3009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By conducting the one-way ANOVA, it is easy to see that the  $p$ -value = 5.936e-10 < 0.05, which tells that there exist at least such a type has an extremely different average weight.

```
library(ggplot2)
library(ggthemes)
library(dplyr)
```

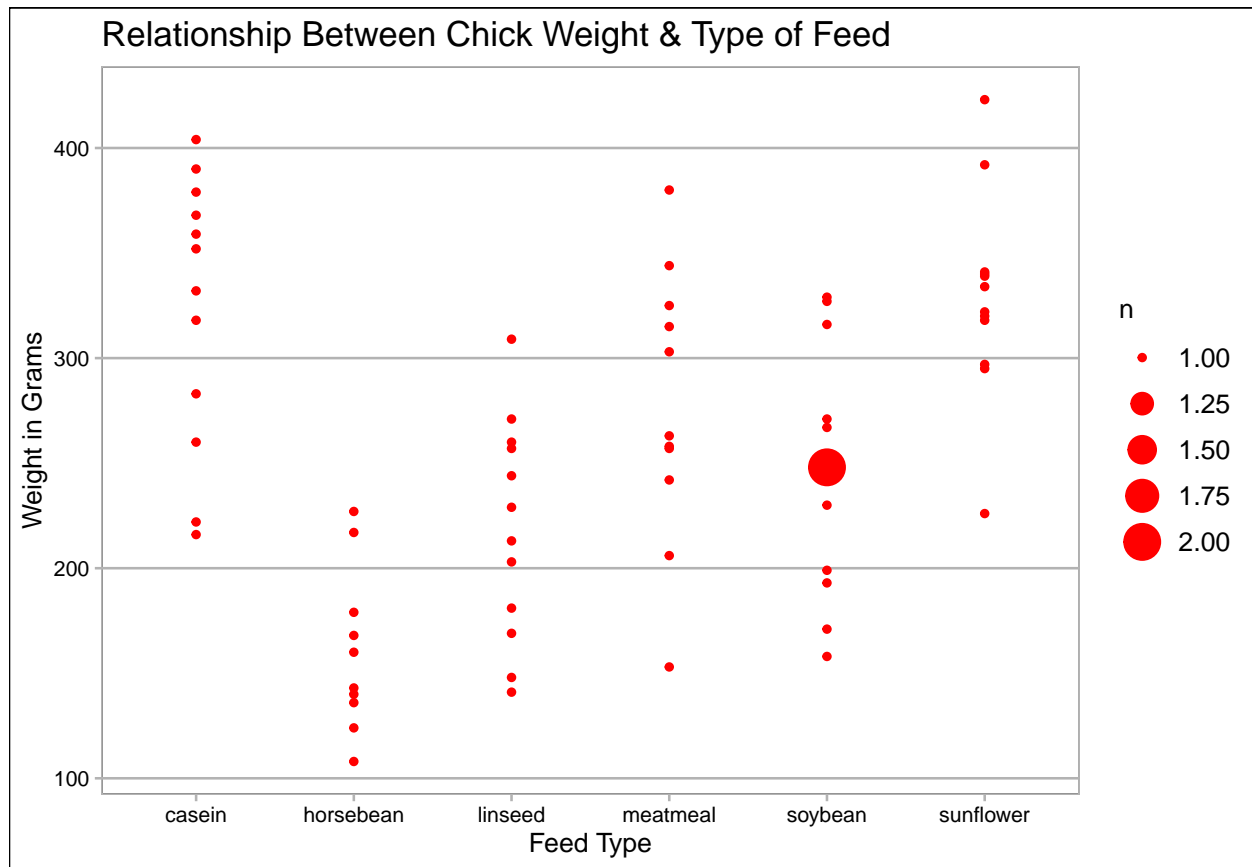
```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
cdata = tbl_df(chickwts)
```

```
## Warning: `tbl_df()` is deprecated as of dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
cdata.wt.feed = cdata %>%
  ggplot(aes(x=feed, y=weight)) +
  geom_count(color="red") +
  labs(title="Relationship Between Chick Weight & Type of Feed",
       x="Feed Type", y="Weight in Grams") +
```

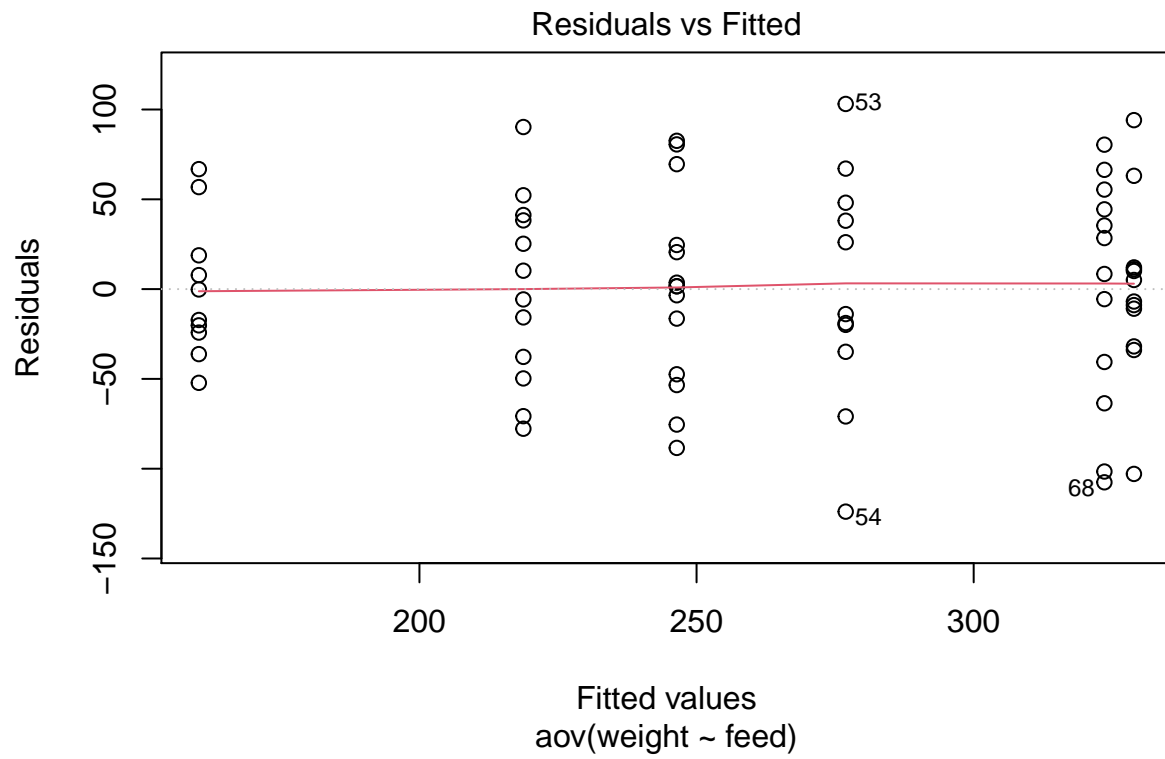
```
theme_calc()
cdata.wt.feed
```

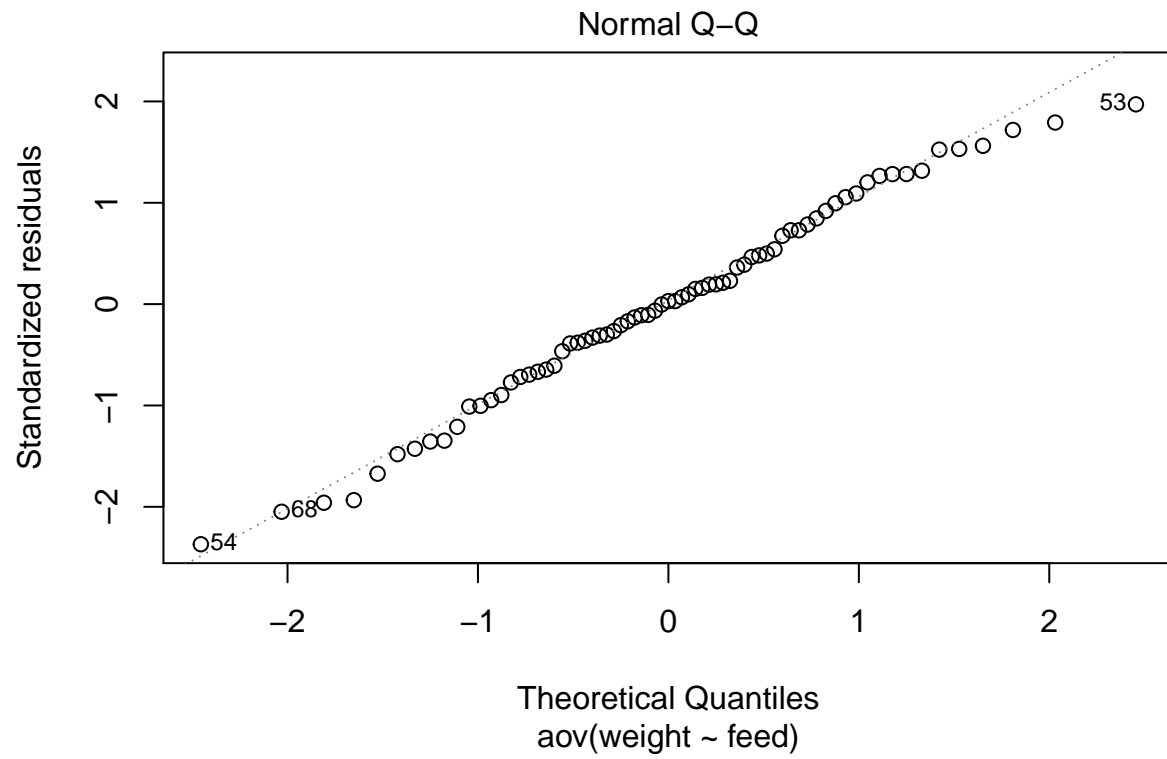


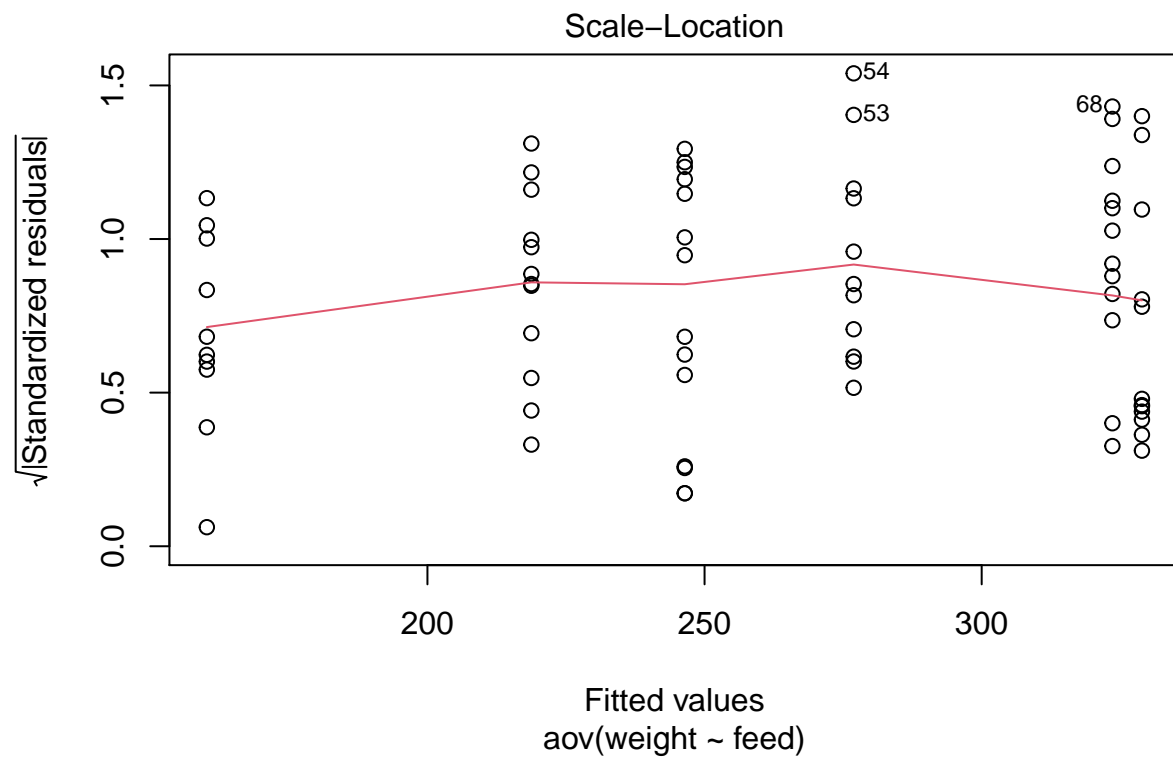
The above plot shows that the casein and sunflower are the best feed supplement.

**c.**

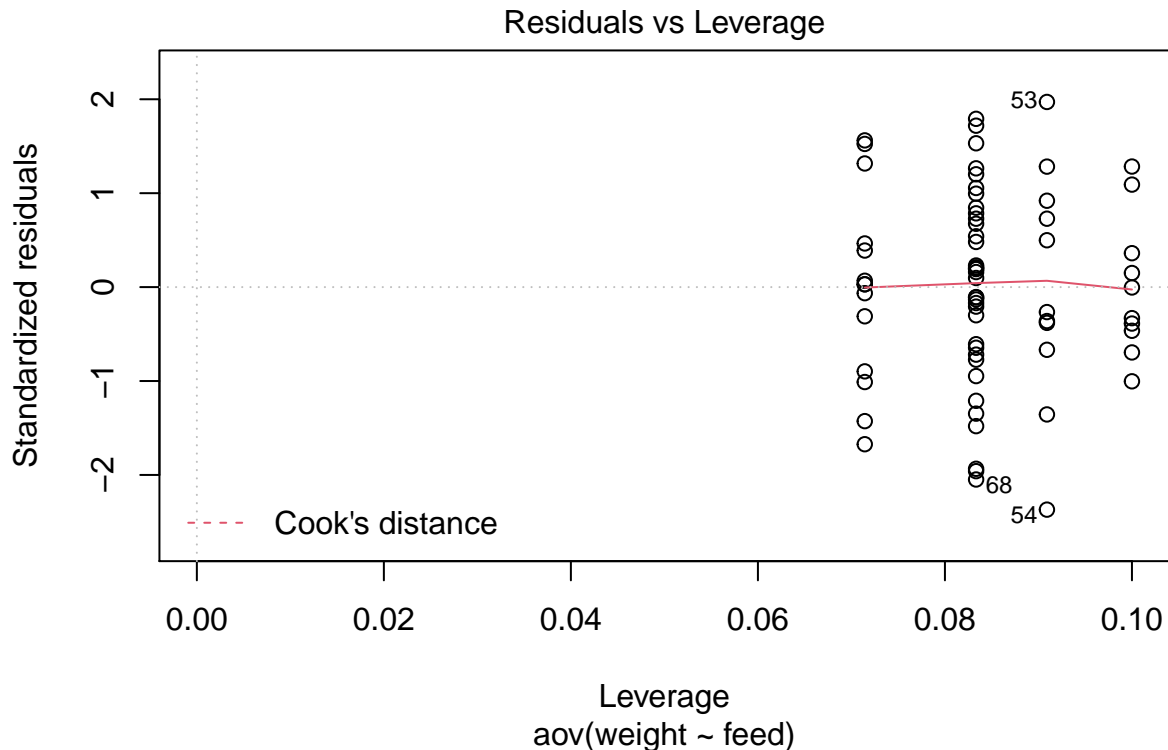
```
caov = aov(weight~feed, data=chickwts)
plot(caov)
```











These plots shows that the ANOVA model assumptions is in line with expectations. In the case like this, we can conclude that values have equal variance.

```
summary(caov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## 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
```

Additionally, since the  $p$ -value  $< 0.05$ , it also shows the normality of the values.

d.

```
kruskal.test(weight ~ feed, data=chickwts)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  weight by feed
## Kruskal-Wallis chi-squared = 37.343, df = 5, p-value = 5.113e-07
```

Kruskal-Wallis test has a  $p$ -value  $= 5.113e-07 < 0.05$ , which means there exist such a feed influenced the weight. And the one-way ANOVA also shows the same conclusion. On the other hand, Kruskal-Wallis test is based on ranks instead of normality as ANOVA. It means that the ANOVA tested the normality of values from means, but Kruskal-Wallis tested on comparison of the ranks of the means.