One-Way ANOVA Test

Andrew Levine

December 2021

```
WL <- read.csv(file="/Users/andrewlevine/Downloads/Statistics II/WeightLoss2.csv")
```

Summary Statistics and Box Plot:

```
str(WL)
```

```
## 'data.frame':    40 obs. of  2 variables:
##  $ program   : chr  "LowCal" "LowCal" "LowCal" "LowCal" ...
##  $ weightloss: int  9 8 6 4 8 9 5 4 0 1 ...
```

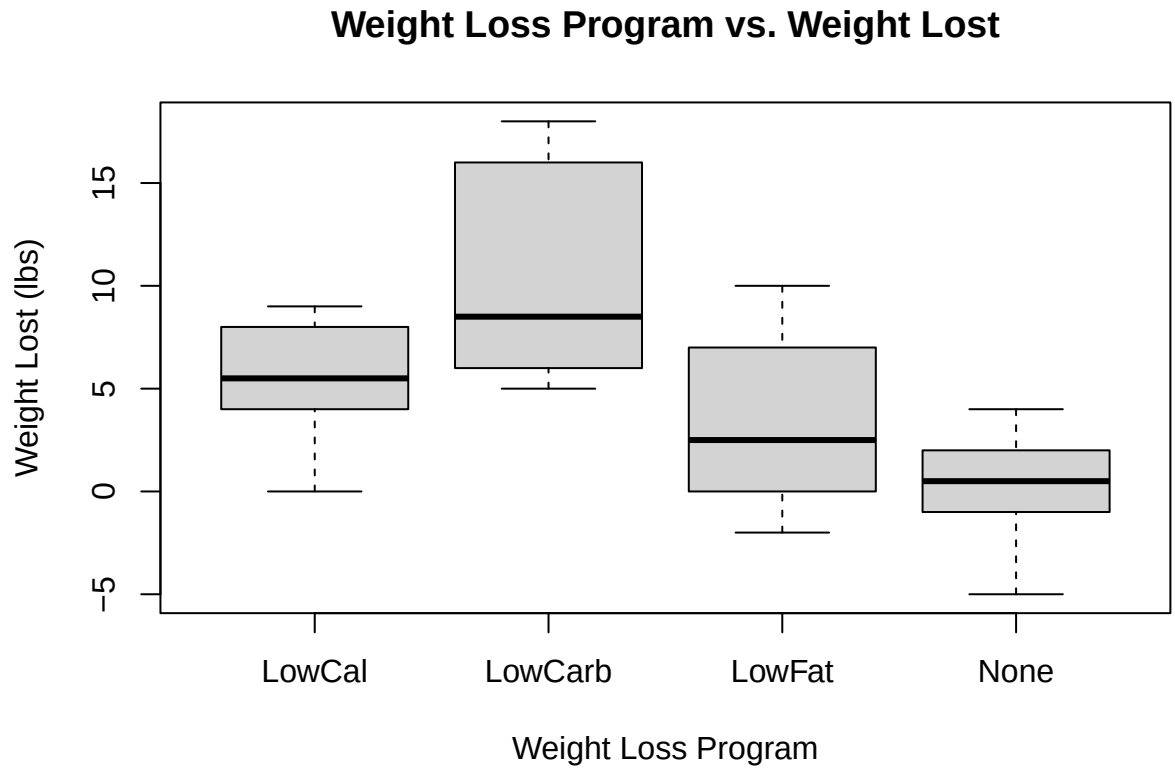
```
WL$program <- factor(WL$program)
library(dplyr)
```

```
##
## 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
```

```
grouped_WL <- group_by(WL, program)
summarize(grouped_WL, mean(weightloss), sd(weightloss), n())
```

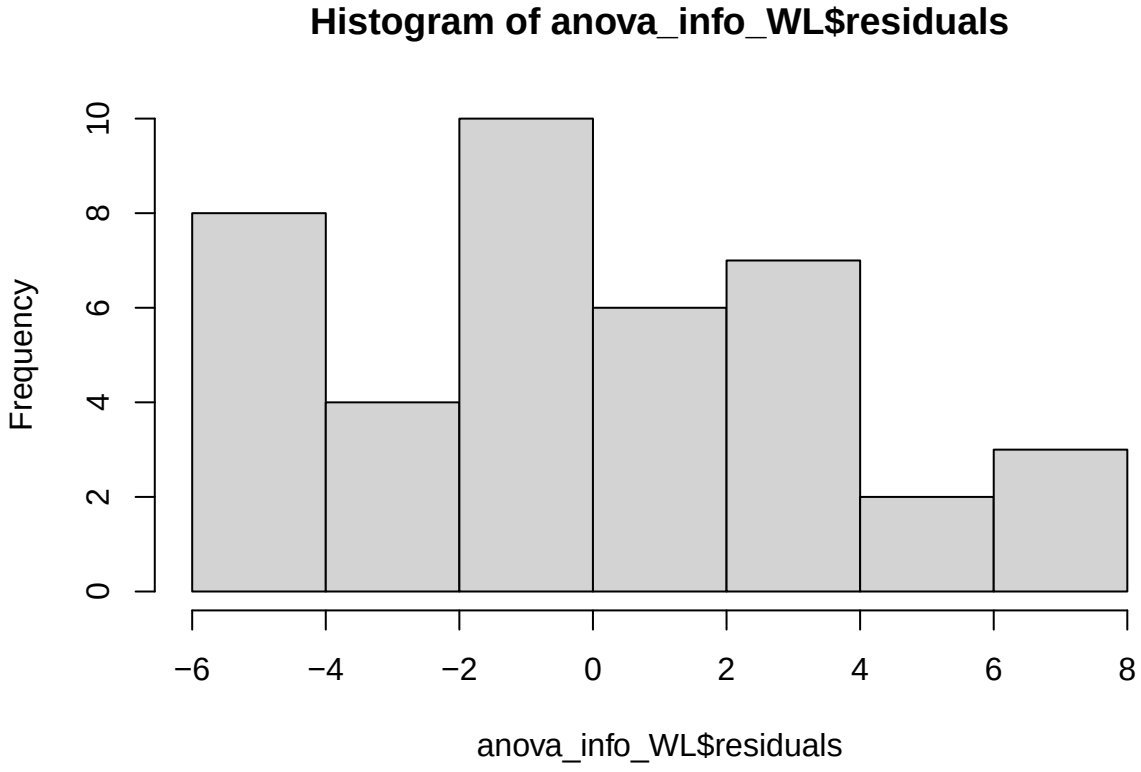
```
## # A tibble: 4 x 4
##   program `mean(weightloss)` `sd(weightloss)` `n()`
##   <fct>                <dbl>            <dbl> <int>
## 1 LowCal                 5.4             3.20    10
## 2 LowCarb               10.5             5.13    10
## 3 LowFat                 3.2             4.18    10
## 4 None                   0.4             2.50    10
```

```
boxplot(weightloss ~ program, data = WL,
        xlab = "Weight Loss Program", ylab = "Weight Lost (lbs)",
        main = "Weight Loss Program vs. Weight Lost"
)
```

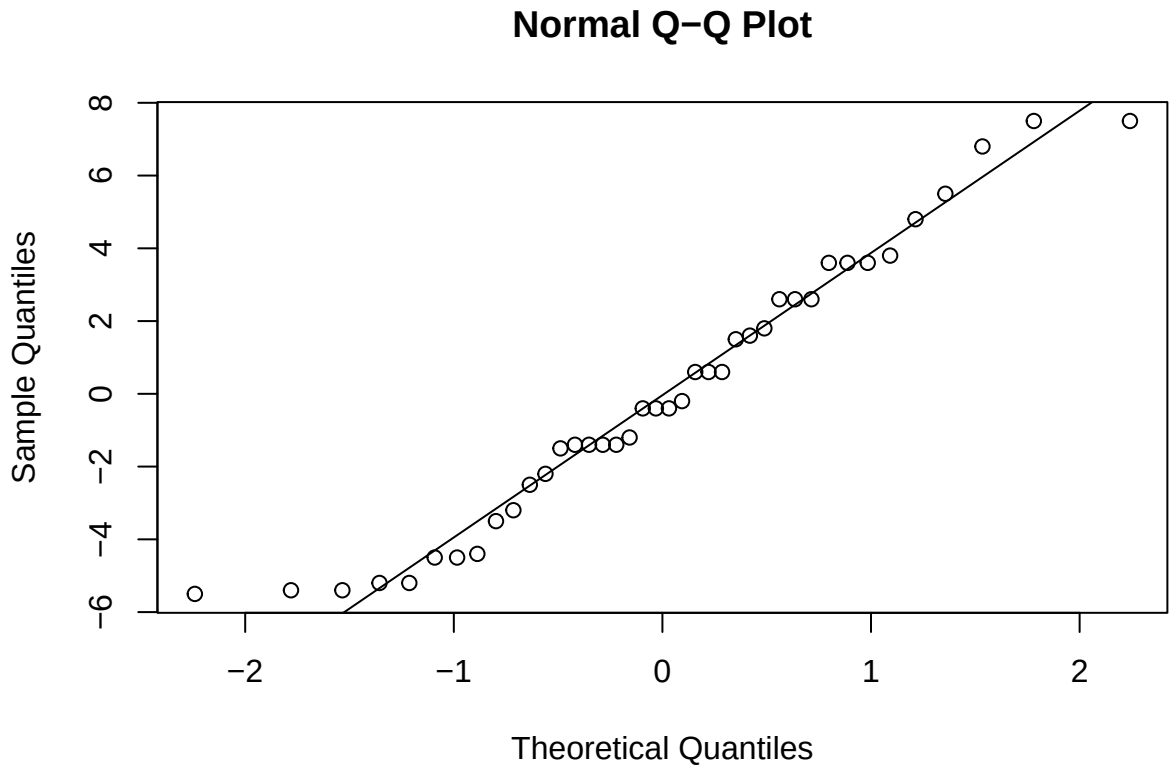


Plots for Linear Assumptions:

```
anova_info_WL <- aov(weightloss ~ program, data = WL)
hist(anova_info_WL$residuals)
```



```
qqnorm(anova_info_WL$residuals)
qqline(anova_info_WL$residuals)
```



Levene Test for Equality of Variances:
5.) Hypotheses and Significance Level:
H_0: $\sigma^2_{LowCarb} = \sigma^2_{LowCal} = \sigma^2_{LowFat} = \sigma^2_{Control}$
H_a: Not all population variances are the same.
α: 0.05

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
leveneTest(weightloss ~ program, data = WL)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3   1.743 0.1756
##       36
```

The Levene test resulted in a p-value of 0.1756, which is larger than our α of 0.05. As a result, it is safe to assume equal variance of errors.

Hypothesis Test:
Parameters:
$\mu_{LowCarb}$ = the mean weight loss (lbs) for people on the low carbohydrate diet
μ_{LowCal} = the mean weight loss (lbs) for people on the low calorie diet
μ_{LowFat} = the mean weight loss (lbs) for people on the low fat diet
μ_{None} = the mean weight loss (lbs) for people in the control group

Hypotheses and Significance Level:
H_0: $\mu_{LowCarb} = \mu_{LowCal} = \mu_{LowFat} = \mu_{None}$
H_a: Not all groups have the same mean weight loss.
α: 0.05

```
summary(anova_info_WL)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## program      3  547.5  182.49    12.1 1.26e-05 ***
## Residuals   36  542.9   15.08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of 1.26 * 10^{-5} is less than the α of 0.05, and so we reject H_0. At the 5% significance level, the average weight loss varies based on weight loss program.
It is appropriate to conduct a post-hoc procedure here, because we did indeed find that mean weight loss does significantly vary based on the program the subject was put in. Since we have evidence suggesting that not all means are the same, we want to determine the ways in which the means differ and by how much. In order to do so, we will run pairwise comparisons with Tukey's HSD test and the Bonferroni correction.

5. Tuykey's HSD Test, ($\alpha = 0.05$):

```
library(DescTools)
```

```
##
## Attaching package: 'DescTools'

## The following object is masked from 'package:car':
##
##     Recode
```

```
PostHocTest(anova_info_WL, conf.level=0.95, method="hsd")
```

```
##
##   Posthoc multiple comparisons of means : Tukey HSD
##     95% family-wise confidence level
##
```

```
## $program
##                   diff      lwr.ci     upr.ci    pval
## LowCarb-LowCal   5.1   0.4226846  9.7773154 0.02812 *
## LowFat-LowCal   -2.2  -6.8773154  2.4773154 0.58949
## None-LowCal     -5.0  -9.6773154 -0.3226846 0.03232 *
## LowFat-LowCarb  -7.3 -11.9773154 -2.6226846 0.00092 ***
## None-LowCarb   -10.1 -14.7773154 -5.4226846 7.1e-06 ***
## None-LowFat     -2.8  -7.4773154  1.8773154 0.38473
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The pairwise p-values for Tukey's HSD test are as follows:
$p\text{-}value_{LowCarb-LowCal} = 0.0281$
$p\text{-}value_{LowFat-LowCal} = 0.5895$
$p\text{-}value_{None-LowCal} = 0.0323$
$p\text{-}value_{LowFat-LowCarb} = 0.0009$
$p\text{-}value_{None-LowCarb} = 7.10 * 10^{-6}$
$p\text{-}value_{None-LowFat} = 0.3847$

Since the p-values for the pairs of LowCarb-LowCal (0.0281), None-LowCal (0.0323), LowFat-LowCarb (0.0009), and None-LowCarb ($7.10 * 10^{-6}$) are all less than the α of 0.05, we can conclude that these pairs each have significant differences in weight loss. The low carbohydrate diet resulted in greater weight loss than all three other programs, as evident by the signs of the difference between the low carbohydrate group and the individual other three. We can also say that the low calorie diet resuled in more weight loss than the control group, judging by the sign of the difference between these two groups. No other significant differences were found between pairs of groups. Due to these findings, we can say that the low carbohydrate diet performed the best, because there were significant differences between this diet and the other three, which the low carbohydrate diet outperformed in terms of average weight lost. These findings also suggest that the control diet performed the worst, due to the fact that the low calorie and low carbohydrate diets both resulted in statistically significant greater weight loss.

The pairwise confidence intervals for Tukey's HSD test are as follows:
$CI_{LowCarb-LowCal} = (0.4227, 9.7773)$
$CI_{LowFat-LowCal} = (-6.8773, 2.4773)$
$CI_{None-LowCal} = (-9.6773, -0.3227)$
$CI_{LowFat-LowCarb} = (-11.9773, -2.6227)$
$CI_{None-LowCarb} = (-14.7773, -5.4227)$
$CI_{None-LowFat} = (-7.4773, 1.8773)$

Since the 95% confidence intervals for the pairs of LowCarb-LowCal (0.4227, 9.7773), None-LowCal (-9.6773, -0.3227), LowFat-LowCarb (-11.9773, -2.6227), and None-LowCarb (-14.7773, -5.4227) do not contain 0, we can conclude that these pairs each have significant differences in weight loss. Since the interval for the pair of LowCarb-LowCal falls above 0, it tells us that the low carbohydrate diet resulted in a higher mean weight loss than the low calorie diet. Since the intervals for the None-LowCal, LowFat-LowCarb, and None-LowCarb pairs are below 0, they tell us that the low calorie diet resulted in higher average weight loss than the control group, and the low carbohydrate diet resulted in higher average weight loss than both the low fat group and the control group. No other significant differences were found. Due to these findings, we can say that the low carbohydrate diet performed the best, because there were significant differences between this diet and the other three, which the low carbohydrate diet outperformed in terms of average weight lost. These

findings also suggest that the control diet performed the worst, due to the fact that the low calorie and low carbohydrate diets both resulted in statistically significant greater weight loss.

Bonferroni Correction, ($\alpha = 0.05$):

```
PostHocTest(anova_info_WL, conf.level=0.95, method="bonferroni")
```

```
##
##   Posthoc multiple comparisons of means : Bonferroni
##     95% family-wise confidence level
##
## $program
##                  diff     lwr.ci    upr.ci    pval
## LowCarb-LowCal   5.1   0.251195  9.948805 0.03452 *
## LowFat-LowCal   -2.2  -7.048805  2.648805 1.00000
## None-LowCal     -5.0  -9.848805 -0.151195 0.04005 *
## LowFat-LowCarb  -7.3 -12.148805 -2.451195 0.00099 ***
## None-LowCarb   -10.1 -14.948805 -5.251195 7.4e-06 ***
## None-LowFat     -2.8  -7.648805  2.048805 0.69384
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a.) The pairwise p-values for the Bonferroni Correction test are as follows:
$p\text{-}value_{LowCarb-LowCal} = 0.0345$
$p\text{-}value_{LowFat-LowCal} = 1.000$
$p\text{-}value_{None-LowCal} = 0.0401$
$p\text{-}value_{LowFat-LowCarb} = 0.0010$
$p\text{-}value_{None-LowCarb} = 7.40 * 10^{-6}$
$p\text{-}value_{None-LowFat} = 0.6938$

Since the p-values for the pairs of LowCarb-LowCal (0.0345), None-LowCal (0.0401), LowFat-LowCarb (0.0010), and None-LowCarb ($7.40 * 10^{-6}$) are all less than the α of 0.05, we can conclude that these pairs each have significant differences in weight loss. The low carbohydrate diet resulted in greater weight loss than all three other programs, as evident by the signs of the difference between the low carbohydrate group and the individual other three. We can also say that the low calorie diet resuled in more weight loss than the control group, judging by the sign of the difference between these two groups. No other significant differences were found between pairs of groups. Due to these findings, we can say that the low carbohydrate diet performed the best, because there were significant differences between this diet and the other three, which the low carbohydrate diet outperformed in terms of average weight lost. These findings also suggest that the control diet performed the worst, due to the fact that the low calorie and low carbohydrate diets both resulted in statistically significant greater weight loss.

The pairwise confidence intervals for the Bonferroni Correction test are as follows:
$CI_{LowCarb-LowCal} =$ (0.2512, 9.9488)
$CI_{LowFat-LowCal} =$ (-7.0488, 2.6488)
$CI_{None-LowCal} =$ (-9.8488, -0.1512)
$CI_{LowFat-LowCarb} =$ (-12.1488, -2.4512)

$CI_{None-LowCarb}$ = (-14.9488, -5.2512)
$CI_{None-LowFat}$ = (-7.6488, 2.0488)

Since the 95% confidence intervals for the pairs of LowCarb-LowCal (0.2512, 9.9488), None-LowCal (-9.8488, -0.1512), LowFat-LowCarb (-12.1488, -2.4512), and None-LowCarb (-14.9488, -5.2512) do not contain 0, we can conclude that these pairs each have significant differences in weight loss. Since the interval for the pair of LowCarb-LowCal falls above 0, it tells us that the low carbohydrate diet resulted in a higher mean weight loss than the low calorie diet. Since the intervals for the None-LowCal, LowFat-LowCarb, and None-LowCarb pairs are below 0, they tell us that the low calorie diet resulted in higher average weight loss than the control group, and the low carbohydrate diet resulted in higher average weight loss than both the low fat group and the control group. No other significant differences were found. Due to these findings, we can say that the low carbohydrate diet performed the best, because there were significant differences between this diet and the other three, which the low carbohydrate diet outperformed in terms of average weight lost. These findings also suggest that the control diet performed the worst, due to the fact that the low calorie and low carbohydrate diets both resulted in statistically significant greater weight loss.