MATH 588

HW3

Md Ismail Hossain

2/14/2022

Question: 4.29

Null Hypothesis: Logging has no effect on the distribution of percentage seedling lose lose between 2004 and 2005.
Alternative Hypothesis: Logging has statistically significant effect on the distribution of percentage seedling lose between 2004 and 2005.

```r
library(Sleuth3)
head(ex0429)
```

```
##     Plot Action Seedlings2004 Seedlings2005 PercentLost
## 1 Plot 1      L           298           164        45.0
## 2 Plot 2      L           471           221        53.1
## 3 Plot 3      L           767           454        40.8
## 4 Plot 4      L           576           141        75.5
## 5 Plot 5      L           407           217        46.7
## 6 Plot 6      L          1534           224        85.4
```

```r
# Median Percent Lost by Action
tapply(ex0429$PercentLost,ex0429$Action,median,na.rm=T)
```

```
##    L    U
## 46.7 18.1
```

```r
# a
# Rank sum test
wilcox.test(PercentLost~Action,conf.int = TRUE,data = ex0429)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  PercentLost by Action
## W = 55, p-value = 0.01154
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  10.8 65.1
## sample estimates:
## difference in location
##                   33.4
```

The p-value (0.01154) is less than 0.05. So, at 5% level of significance we can reject the null hypothesis and conclude than logging has significant impact on the distribution of percentage of seedlings between 2004 and 2005.

The 95% confidence interval on the difference in median is $(10.8, 65.1)$

```r
#b
t.test(PercentLost~Action,var.equal=TRUE,data = ex0429)
```

```
##
##  Two Sample t-test
##
## data:  PercentLost by Action
## t = 3.1557, df = 14, p-value = 0.007011
## alternative hypothesis: true difference in means between group L and group U is not equal to 0
## 95 percent confidence interval:
##  12.20708 64.00245
## sample estimates:
```

```
## mean in group L mean in group U
##        54.83333        16.72857
```

Comparing the rank-sum test results and independent sample t-test results, both giving us the same decision of rejecting the null hypothesis. But the 95% CI interval on the difference in the medians using rank-sum little smaller than the t-tools CI with same level of significance.

Question 5.15

```
# a
sample_var = var(case0502$Percent)

# b
total_ss = sample_var*45
total_ss
```

```
## [1] 3791.526
```

```
# c

sp = ((9-1)*5.03^2+(5-1)*11.94^2+(6-1)*6.58^2+(9-1)*4.59^2+(2-1)*3.818^2+(6-1)*9.010^2+(9-1)*5.969^2)/(
residual_ss = sp*(46-7)
residual_ss
```

```
## [1] 1863.198
```

```
ms_resid = residual_ss/(46-7)
ms_resid
```

```
## [1] 47.7743
```

```
# SS Between groups
ss_btwn = total_ss-residual_ss
ss_btwn
```

```
## [1] 1928.328
```

```
ms_btwn = ss_btwn/6
ms_btwn
```

```
## [1] 321.3881
```

```
# F-statistic
f_statistic = ms_btwn/ms_resid
f_statistic
```

```
## [1] 6.727217
```

```r
#d
pf(f_statistic,6,39, lower.tail = FALSE)
```

```
## [1] 6.023828e-05
```

So, we are getting exactly similar results like in Disply 5.10

Question: 5.21

Null Hypothesis: All population variance are equal.
Alternative Hypothesis: Populaiton variance are unequal.

```r
#install.packages("car")
library(car)
# Levene's test with one independent variable
leveneTest(Percent ~ Judge, data = case0502)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  6  1.2625 0.2969
##       39
```

At 5% level of significance we can not reject the null hypothesis because p-value (0.2969) is higher than 0.05 and conclude that population variance are equal.

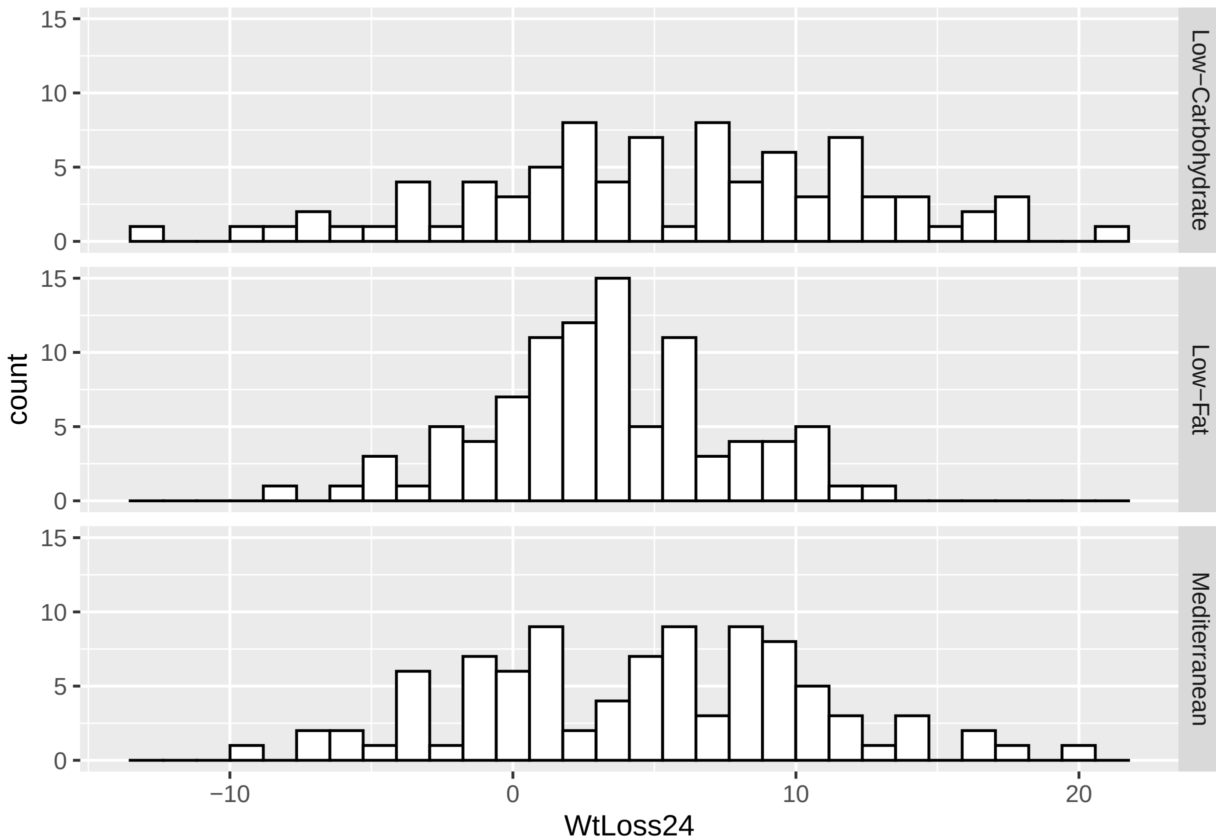
Question: 6.23

```r
head(ex0623)
```

```
##   Subject   Group WtLoss24
## 1       1 Low-Fat      2.2
## 2       2 Low-Fat     -4.8
## 3       3 Low-Fat      2.9
## 4       4 Low-Fat      4.4
## 5       5 Low-Fat      9.6
## 6       6 Low-Fat      6.3
```

```r
# install.packages("ggplot2")
library(ggplot2)

ggplot(ex0623, aes(x = WtLoss24)) +
  geom_histogram(fill = "white", colour = "black") +
  facet_grid(Group ~ .)
```



The histogram suggest that data does not suffer severally from non normality. So, we can use one way anova to do the comparison among groups.

```r
# Compute the analysis of variance
anova_oneway <- aov(WtLoss24 ~ Group, data = ex0623)
# Summary of the analysis
summary(anova_oneway)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Group         2    217  108.43   3.236 0.0409 *
## Residuals   269   9014   33.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value 0.0409 is less than 0.05, so at 5% level of significance we can conclude that there is statistically significant difference in terms of mean weight loss between at least one pair of the diet plan groups.

```r
# Multiple test:
TukeyHSD(anova_oneway)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = WtLoss24 ~ Group, data = ex0623)
##
## $Group
##                                        diff       lwr        upr     p adj
## Low-Fat-Low-Carbohydrate       -2.1828035 -4.224771 -0.1408361 0.0329364
```

```
## Mediterranean-Low-Carbohydrate -0.8849083 -2.932082  1.1622656 0.5656813
## Mediterranean-Low-Fat           1.2978952 -0.697418  3.2932084 0.2771180
```

Turkeys HSD test indicates that there is statistically significant difference between Low-Fat Vs Low-Carbohydrate diet groups.

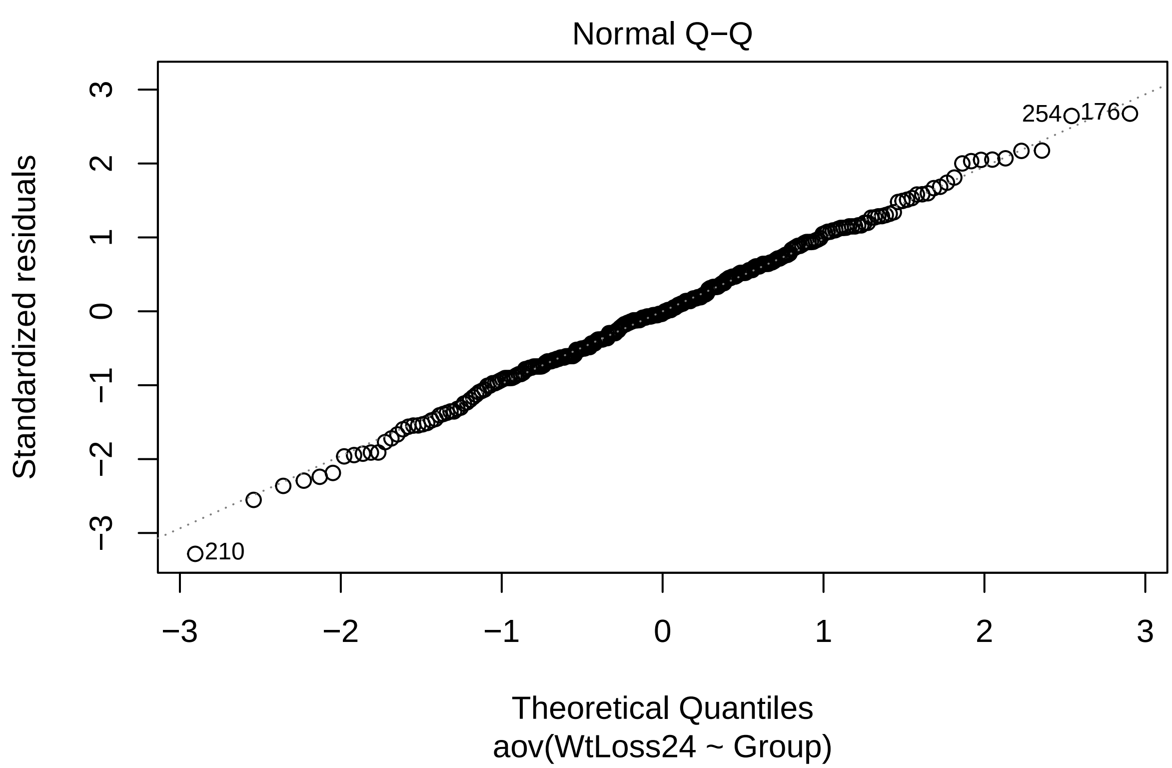
```
aggregate(WtLoss24 ~ Group, data = ex0623, FUN = mean)
```

```
##              Group WtLoss24
## 1 Low-Carbohydrate 5.487059
## 2          Low-Fat 3.304255
## 3    Mediterranean 4.602151
```

And among all the diet plans, Low-Fat diet have the lowest mean weight loss. So, if losing weight is the goal for an individual then Low-Fat diet would be the winner in that case.

Checking ANOVA assumptions

```
# Normality
plot(anova_oneway, 2)
```



The normality assumption does not looks violated.

```
# Equal variance
leveneTest(WtLoss24 ~ Group, data = ex0623)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group   2  12.957 4.243e-06 ***
```

```
##       269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This test of homogeneity suggest that, at 5% level of significance we can conclude that the groups do not have equal variance. Equality of variance is one of the assumption settled before carry out the anova. The alternative is to use non parametric approach to test the hypothesis or we could allow this inequality of the variance among groups, because in practical situation equality among groups does not found in most case.

Non-parametric approach:

We are getting exactly similar types of results like ANOVA even after using non-parametric approach.

```
kruskal.test(WtLoss24 ~ Group, data = ex0623)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  WtLoss24 by Group
## Kruskal-Wallis chi-squared = 6.6058, df = 2, p-value = 0.03678
```

```
pairwise.wilcox.test(ex0623$WtLoss24,ex0623$Group,
                 p.adjust.method = "BH")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  ex0623$WtLoss24 and ex0623$Group
##
##               Low-Carbohydrate Low-Fat
## Low-Fat       0.031            -
## Mediterranean 0.300            0.201
##
## P value adjustment method: BH
```