

HW5

Hanao Li

October 5, 2019

Question 1

a)

```
if(!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
p_load(car)
```

```
data <- subset(ChickWeight, Time == 18)
anova18 <- aov(data=data, weight ~ Diet)
anova18
```

```
## Call:
##   aov(formula = weight ~ Diet, data = data)
##
## Terms:
##              Diet Residuals
## Sum of Squares   36690.44 114840.84
## Deg. of Freedom      3      43
##
## Residual standard error: 51.67898
## Estimated effects may be unbalanced
```

```
summary(anova18)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Diet           3  36690   12230   4.579 0.0072 **
## Residuals     43 114841    2671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the results, we could see that our p-value is less than 0.05, so we could reject the null hypothesis and conclude there is a significant difference in the mean weights of the four groups on Day 18.

b)

```

birth <- subset(ChickWeight, Time == 0)
data[, 'birthweight'] <- birth$weight[match(data$Chick, birth$Chick)]

anova18_birth = aov(data=data, weight ~ birthweight + Diet)
anova18_birth

```

```

## Call:
## aov(formula = weight ~ birthweight + Diet, data = data)
##
## Terms:
##          birthweight      Diet Residuals
## Sum of Squares    15728.23  27190.80 108612.25
## Deg. of Freedom         1         3         42
##
## Residual standard error: 50.85279
## Estimated effects may be unbalanced

```

```
summary(anova18_birth)
```

```

##          Df Sum Sq Mean Sq F value Pr(>F)
## birthweight  1  15728   15728   6.082 0.0178 *
## Diet         3  27191    9064   3.505 0.0234 *
## Residuals   42 108612    2586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

From the results, we could see that both our p-values are less than 0.05, so we could reject the null hypothesis and conclude that there is a significant difference in the mean weights of the four groups on Day 18 after adjusting for Birth Weight.

c)

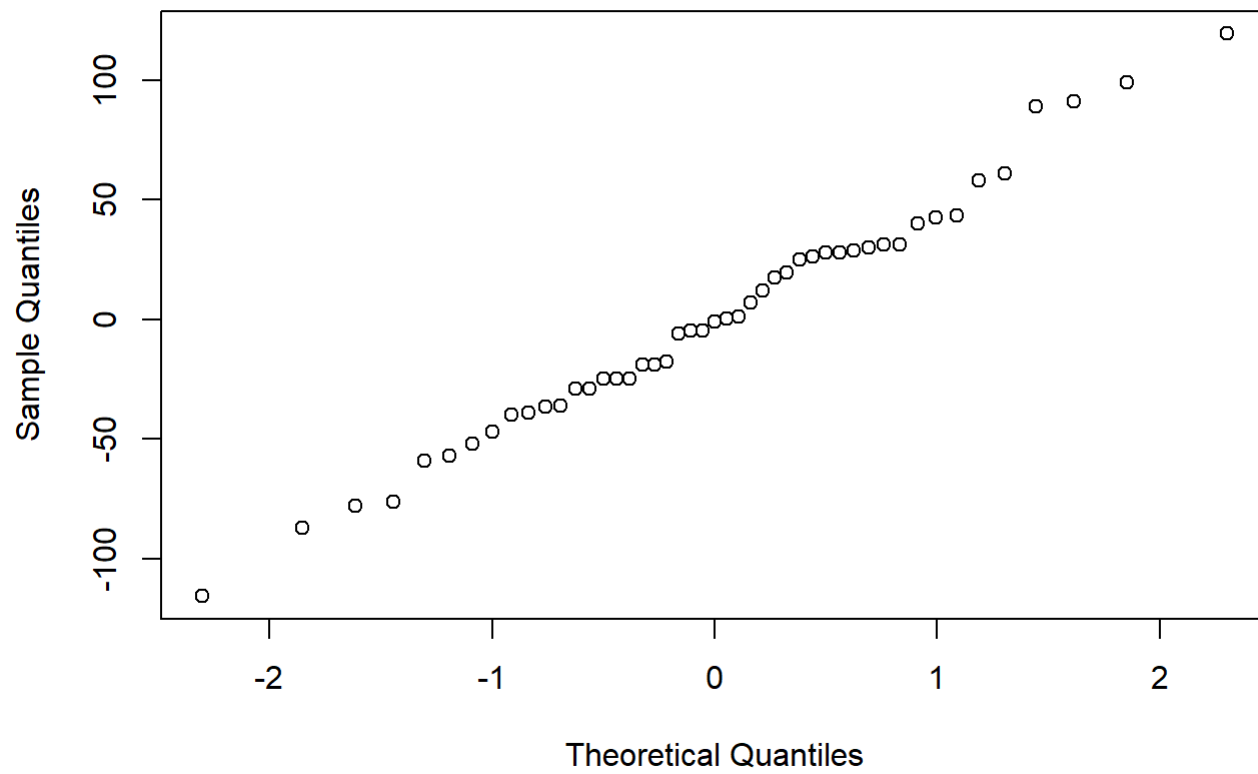
```

# We assume that the data are IID normal and they have constant variance.

qqnorm(unlist(anova18['residuals']))

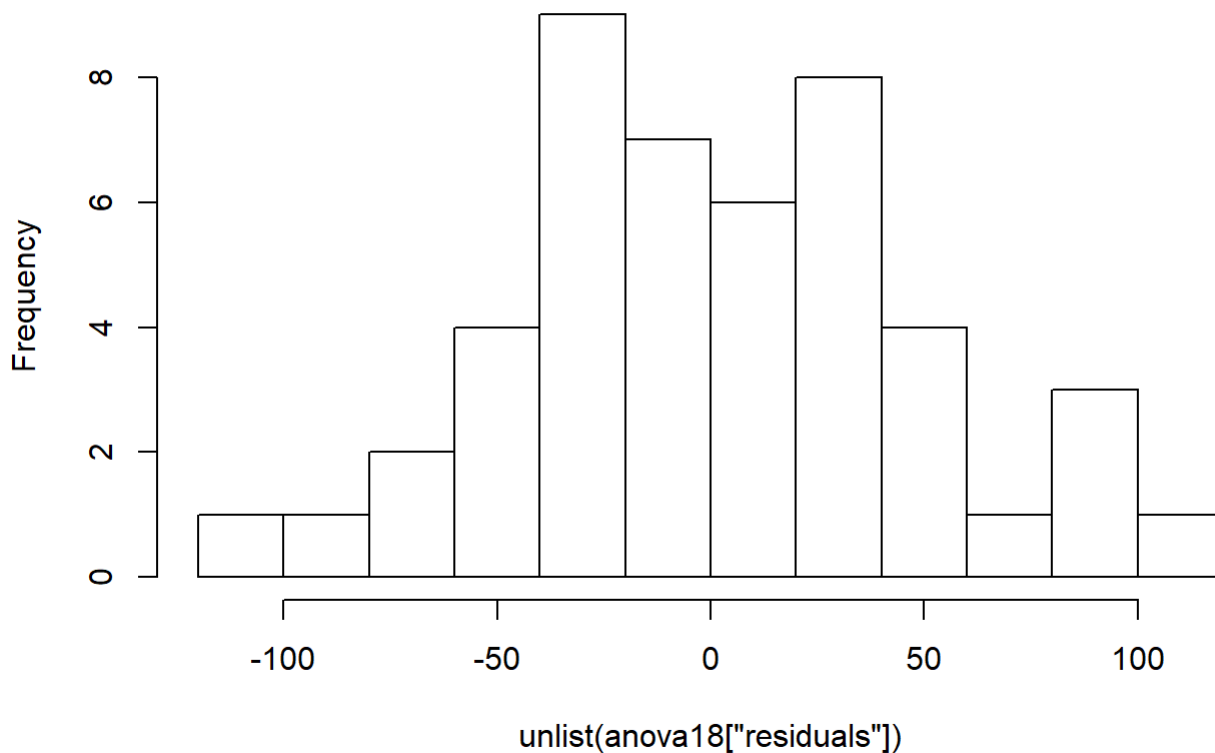
```

Normal Q-Q Plot



```
hist(unlist(anova18['residuals']), breaks=10, freq=T)
```

Histogram of `unlist(anova18["residuals"])`



```
shapiro.test(unlist(anova18['residuals']))
```

```
##
## Shapiro-Wilk normality test
##
## data:  unlist(anova18["residuals"])
## W = 0.98863, p-value = 0.9242
```

```
kruskal.test(weight ~ Diet, data=data)
```

```
##
## Kruskal-Wallis rank sum test
##
## data:  weight by Diet
## Kruskal-Wallis chi-squared = 10.623, df = 3, p-value = 0.01395
```

From the plots we could see that the data seems to be following a normal distribution and shapiro's test proved this since the p-value is larger than 0.05, we do not reject the null and conclude data satisfy normal distribution. And from the kruskal's test, we could see that p-value is less than 0.05, so we reject the null and conclude the weight data in the four diet group are not from the same distribution.

```
bartlett.test(x=data$weight, g=data$Diet)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: data$weight and data$Diet
## Bartlett's K-squared = 3.576, df = 3, p-value = 0.311
```

```
leveneTest(y=data$weight, group=as.factor(data$Diet))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3   1.059 0.3764
##      43
```

Since both bartlett's test and Levene's test have p-values larger than 0.05, we do not reject the null and conclude that the data satisfies the constant variance assumption.

```
# We also need to check parallelism
para <- aov(weight ~ birthweight * Diet, data=data)
summary(para)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## birthweight    1  15728   15728   6.134 0.0177 *
## Diet           3   27191    9064   3.535 0.0234 *
## birthweight:Diet 3    8610    2870   1.119 0.3530
## Residuals     39 100002    2564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the result we could see that the p-value is 0.35 larger than 0.05 so we do not reject the null and conclude that there is no interaction between birthweight and Diet. Therefore, marginal inference on these two variables is not needed.

Question 2

a)

```
pairwise.t.test(data$weight, data$Diet, p.adj="bonferroni")
```

```
##
## Pairwise comparisons using t tests with pooled SD
##
## data: data$weight and data$Diet
##
##      1      2      3
## 2 1.0000 -      -
## 3 0.0049 0.3358 -
## 4 0.2313 1.0000 1.0000
##
## P value adjustment method: bonferroni
```

b)

```
TukeyHSD(anova18)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = weight ~ Diet, data = data)
##
## $Diet
##          diff          lwr          upr      p adj
## 2-1  28.75882 -26.28093  83.79857 0.5084630
## 3-1  74.15882  19.11907 129.19857 0.0043760
## 4-1  43.95882 -11.08093  98.99857 0.1586149
## 3-2  45.40000 -16.36380 107.16380 0.2173462
## 4-2  15.20000 -46.56380  76.96380 0.9122642
## 4-3 -30.20000 -91.96380  31.56380 0.5637852
```

From the results, we could see that bonferroni method generally results higher p-value. But if we set the alpha value equal to 0.05, we will have the same results from these two methods. We will reject the null for diet group 3-1 and not reject other groups.

Question 3

```
kruskal.test(data$weight, data$Diet)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: data$weight and data$Diet
## Kruskal-Wallis chi-squared = 10.623, df = 3, p-value = 0.01395
```

From the kruskal's test, we could see that p-value is less than 0.05, so we reject the null and conclude the weight data in the four diet group are not from the same distribution.