GR5291 - HW5

Mathieu Sauterey - UNI: mjs2364

October 10, 2018

Consider the ChickWeight data in R. The body weights of the chicks were measured at birth (i.e., time = 0) and every second day thereafter until day 20. They were also measured on day 21. There were four groups of chicks on different protein diets.

1. Determine whether there is a significant difference in the mean weights of the four groups on Day 21:

a) Without adjusting for Birth Weight

```r
# Loads the data
attach(ChickWeight)

# Subset on day 21
day_21 = subset(ChickWeight, Time == 21)

# Performs One-way ANOVA which assumes normality and equal variance across groups
fit <- aov(weight ~ Diet, data = day_21)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet         3  57164   19055   4.655 0.00686 **
## Residuals   41 167839    4094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We recall that for One-way ANOVA:

- Null hypothesis: the mean weights of the different diet groups are the same
- Alternative hypothesis: At least one sample mean weight is not equal to the others.

P-value < 0.05 therefore we reject the null hypothesis and conclude that on Day 21 at least one sample mean is not equal to the others.

b) Adjusting for Birth Weight. Give the LS Means (i.e., adjusted for Birth Weight)

```r
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 3.4.4
```

```
## The 'lsmeans' package is being deprecated.
## Users are encouraged to switch to 'emmeans'.
## See help('transition') for more information, including how
## to convert 'lsmeans' objects and scripts to work with 'emmeans'.
```

```r
# Finds all chicken with a record on Day 21
chicken_nb = day_21$Chick

# Creates new vector of Day 0 weights (baseline) for those chicken still there on Day 21
day_21$weight_0 = ChickWeight$weight[ChickWeight$Chick %in% chicken_nb & ChickWeight$Time == 0]

# Performs ANCOVA
fit_adjusted <- aov(weight ~ weight_0+Diet, data = day_21)
summary(fit_adjusted)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## weight_0     1  20538   20538   5.112 0.0293 *
## Diet         3  43763   14588   3.631 0.0208 *
## Residuals   40 160703    4018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Calculates LS Means
lsmeans(fit_adjusted, ~ Diet)
```

```
##  Diet   lsmean       SE df lower.CL upper.CL
##  1    183.6429 16.45134 40 150.3935 216.8923
##  2    210.3423 20.30879 40 169.2967 251.3878
##  3    267.1307 20.18443 40 226.3365 307.9250
##  4    236.4427 21.18748 40 193.6212 279.2642
##
## Confidence level used: 0.95
```

P-value < 0.05 therefore we reject the null hypothesis and conclude that on Day 21 after adjusted for Birth Weights at least one sample mean is not equal to the others.

LS means are:

Diet 1: $\mu_1 = 183.6$ Diet 2: $\mu_2 = 210.3$ Diet 3: $\mu_3 = 267.1$ Diet 4: $\mu_4 = 236.4$

2. For 1a), perform pairwise comparisons among the 4 groups using each of the following, and comment on the results

- Bonferroni method

- Tukey method

```r
# Bonferroni method
pairwise.t.test(weight, Diet, p.adj = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  weight and Diet
##
##   1       2       3
## 2 0.06838 -       -
```

```
## 3 2.5e-06 0.14077 -
## 4 0.00026 0.95977 1.00000
##
## P value adjustment method: bonferroni
```

```r
# Tukey method
TukeyHSD(fit)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight ~ Diet, data = day_21)
##
## $Diet
##          diff        lwr       upr     p adj
## 2-1  36.95000  -32.11064 106.01064 0.4868095
## 3-1  92.55000   23.48936 161.61064 0.0046959
## 4-1  60.80556  -10.57710 132.18821 0.1192661
## 3-2  55.60000  -21.01591 132.21591 0.2263918
## 4-2  23.85556  -54.85981 102.57092 0.8486781
## 4-3 -31.74444 -110.45981  46.97092 0.7036249
```

The pairwise comparison among the 4 groups yields the following results:

- The Bonferroni method shows that at the 95% confidence level, Diet 3 and Diet 4 have a different mean weight than Diet 1
- The Tukey method shows that at the 95% confidence level, only Diet 3 has a different mean weight than Diet 1.

3. Repeat 1a) using the Kruskal-Wallis test

```r
# Performs Kruskal-Wallis test which assume distribution are same across groups
kruskal.test(day_21$weight, day_21$Diet)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  day_21$weight and day_21$Diet
## Kruskal-Wallis chi-squared = 10.585, df = 3, p-value = 0.0142
```

We recall that for Kruskal-Wallis test:

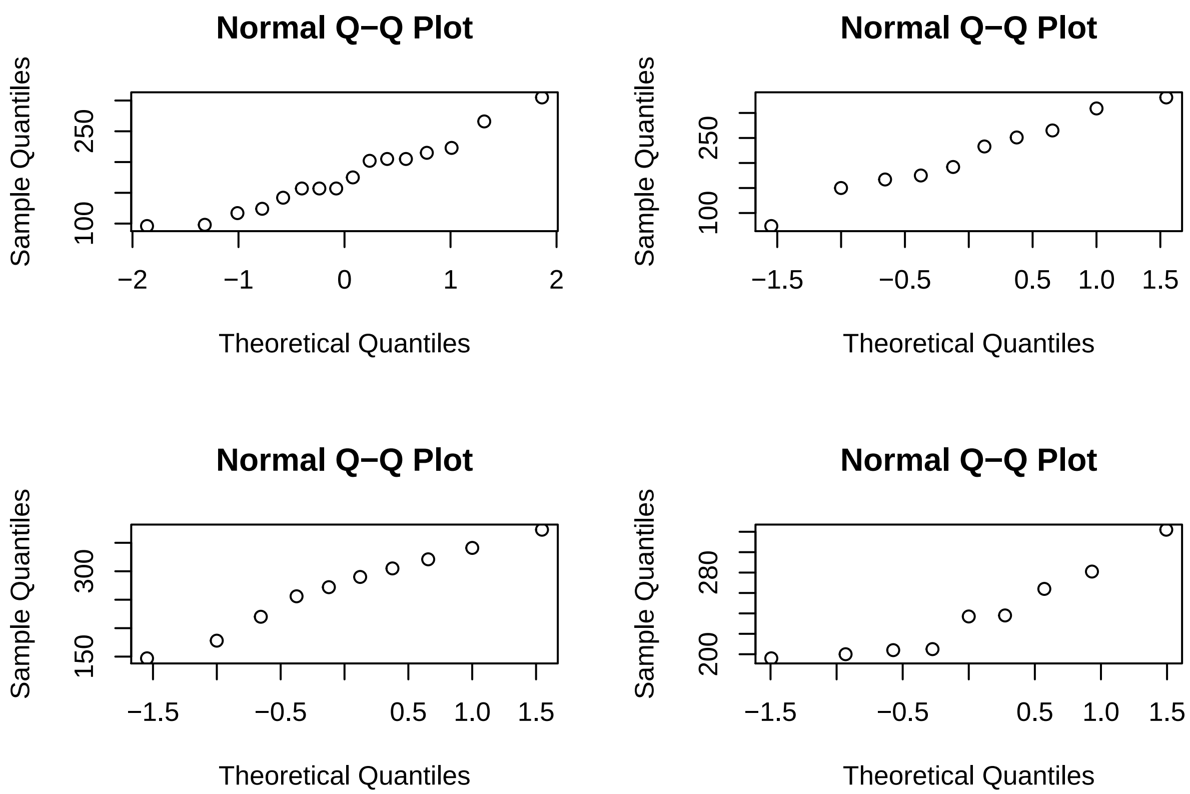
- Null hypothesis: the mean weights of the different diet groups are the same
- Alternative hypothesis: At least one sample mean weight is not equal to the others.

P-value < 0.05 therefore we reject the null hypothesis and conclude that on Day 21 there are significant differences between the average weights of the 4 treatment groups. In other words, we obtain the same conclusion as with One-Way ANOVA.

4. For 1(a) and (b) above, check the validity of your assumptions, including parallelism. Suggest measures that you would take if the assumptions are not satisfied.

```r
# First we check the normality assumption for each group (i.e diet)

# 1. using QQplots
par(mfrow=c(2,2))
qqnorm(day_21$weight[day_21$Diet == 1])
qqnorm(day_21$weight[day_21$Diet == 2])
qqnorm(day_21$weight[day_21$Diet == 3])
qqnorm(day_21$weight[day_21$Diet == 4])
```




```r
# 2. using the Shapiro-Wilk test for normality
shapiro.test(day_21$weight[day_21$Diet == 1])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  day_21$weight[day_21$Diet == 1]
## W = 0.95602, p-value = 0.5905
```

```r
shapiro.test(day_21$weight[day_21$Diet == 2])
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  day_21$weight[day_21$Diet == 2]
## W = 0.97725, p-value = 0.9488
```

```
shapiro.test(day_21$weight[day_21$Diet == 3])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  day_21$weight[day_21$Diet == 3]
## W = 0.97045, p-value = 0.895
```

```
shapiro.test(day_21$weight[day_21$Diet == 4])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  day_21$weight[day_21$Diet == 4]
## W = 0.88694, p-value = 0.1855
```

First we checked the normality assumption underlying the One-Way ANOVA:

The QQplots display an approximately straight line which suggests normality. Moreover, the Shapiro-Wilk test, which has null-hypothesis of normal distribution, fails to reject the null hypothesis at the 95% confidence interval which also suggests normality. Therefore, we conclude that the normality assumption is not violated.

```
# Second we check the assumption of equal variance for each group (i.e diet)

# 1. using Bartlett's test as the groups are normally distributed
bartlett.test(weight ~ Diet, data = day_21)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  weight by Diet
## Bartlett's K-squared = 3.0524, df = 3, p-value = 0.3836
```

Second we check the assumption of equal variance between groups:

To do so, we use the Bartlett's test because it is highly dependent on the normality assumption which we just verified. The null hypothesis of Bartlett's test is equal variance for each group. The p-value > 0.05 therefore there is no evidence to suggest that the variance in weight is statistically significantly different for the four treatment groups. Therefore, we conclude that the assumption of equal variance is not violated.

```
# Third we check the parallelism assumption

# Test for parallelism
summary(aov(weight ~ weight_0*Diet, data = day_21))
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## weight_0        1  20538   20538   5.734 0.0218 *
## Diet            3  43763   14588   4.073 0.0135 *
## weight_0:Diet   3  28185    9395   2.623 0.0649 .
## Residuals      37 132517    3582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Third we check the assumption of parallelism:

The interraction term between Birth Weight and Diet is not statistically significant at the 95% confidence level (P-value > 0.05) therefore the null hypothesis of parallelism assumption is not violated.