5291 hw5

Yijin Wang

10/14/2020

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.

Determine whether there is a significant difference in the mean weights of the four groups on Day 21: a)Without adjusting for Birth Weight

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

```r
#a) Without adjusting for Birth Weight
data(ChickWeight)
day21<-subset(ChickWeight, Time == 21)
anova21<-aov(weight ~ Diet, ChickWeight)
summary(anova21)
```

```
##              Df  Sum Sq Mean Sq F value   Pr(>F)    
## Diet          3  155863   51954   10.81 6.43e-07 ***
## Residuals   574 2758693    4806                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Since p-value < 0.05, we reject the null hypothese.
#Thus, there is a significant difference in the mean weights of 4 groups
#on day21.

#b)Adjusting for Birth Weight.
day0<-subset(ChickWeight, Time == 0)
day21[,"birthweight"]<-day0$weight[match(day21$Chick, day0$Chick)]
day21[,"weight_diff"]<-day21$weight-day21$birthweight

anova21adjust<-aov(weight_diff ~ Diet, day21)
summary(anova21adjust)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)   
## Diet         3  58130   19377   4.698 0.00655 **
## Residuals   41 169091    4124                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#LS means
#install.packages("emmeans")
library("emmeans")
lsmeans(anova21adjust,"Diet")
```

```
##  Diet lsmean   SE df lower.CL upper.CL
##  1       136 16.1 41      104      169
```

```
##  2       174 20.3 41      133      215
##  3       230 20.3 41      188      271
##  4       198 21.4 41      154      241
##
## Confidence level used: 0.95
```

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

a) Bonferroni method

b) Tukey method

```
#a) Bonferroni method
day21$Diet<-factor(day21$Diet)
pairwise.t.test(day21$weight,day21$Diet, p.adj="bonf")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  day21$weight and day21$Diet
##
##   1      2      3
## 2 0.9573 -      -
## 3 0.0053 0.3533 -
## 4 0.1669 1.0000 1.0000
##
## P value adjustment method: bonferroni
```

```
#b) Tukey method
TukeyHSD(anova21)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight ~ Diet, data = ChickWeight)
##
## $Diet
##          diff         lwr      upr     p adj
## 2-1 19.971212  -0.2998092 40.24223 0.0552271
## 3-1 40.304545  20.0335241 60.57557 0.0000025
## 4-1 32.617257  12.2353820 52.99913 0.0002501
## 3-2 20.333333  -2.7268370 43.39350 0.1058474
## 4-2 12.646045 -10.5116315 35.80372 0.4954239
## 4-3 -7.687288 -30.8449649 15.47039 0.8277810
```

We conclude from the results that there are large differences between groups especially group 1 and group 3. The p-value is pretty small and difference is large.

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

```
kruskal.test(weight~Diet, day21)
```

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

Since the p-value is $0.0142 < 0.05$, we reject the null hypothese. Thus, the weight in 4 groups are nonidentical distribution.