5291 hw6

Yijin Wang

10/30/2020

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.

1. Perform ANCOVA, adjusting for baseline, to determine whether there is a significant difference in the mean weights of the four groups using the measurements separately at each timepoint: Day 10, 18, and 20.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
day21<-ChickWeight%>%filter(Time==21)
#day21 <- day21 %>%
 #   mutate(Chick = as.numeric(Chick))

day0<-ChickWeight%>%filter(Time==0)
#day0 <- day0 %>%
 #   mutate(Chick = as.numeric(Chick))

day10<-ChickWeight%>%filter(Time==10)
#day10 <- day10 %>%
#    mutate(Chick = as.numeric(Chick))

day18<-ChickWeight%>%filter(Time==18)
#day18 <- day18 %>%
 #   mutate(Chick = as.numeric(Chick))

day20<-ChickWeight%>%filter(Time==20)
#day20 <- day20 %>%
#    mutate(Chick = as.numeric(Chick))

day21_0<-left_join(day21, day0, by="Chick")%>%
  mutate(weight=weight.x, base=weight.y)%>%
  select(c(weight, base, Time.x, Chick, Diet.x))
```

```r
day20_0<-left_join(day20, day0, by="Chick")%>%
  mutate(weight=weight.x, base=weight.y)%>%
  select(c(weight, base, Time.x, Chick, Diet.x))

day18_0<-left_join(day18, day0, by="Chick")%>%
  mutate(weight=weight.x, base=weight.y)%>%
  select(c(weight, base, Time.x, Chick, Diet.x))

day10_0<-left_join(day10, day0, by="Chick")%>%
  mutate(weight=weight.x, base=weight.y)%>%
  select(c(weight, base, Time.x, Chick, Diet.x))


library(car)
```

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

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

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

```
## The following object is masked from 'package:purrr':
##
##     some
```

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

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

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

#ANCOVA
#Day10
ancova10<-aov(weight~Diet.x+base, data=day10_0)
summary(ancova10)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet.x       3   8314  2771.4   6.336 0.00115 **
## base         1     57    57.3   0.131 0.71908
## Residuals   44  19247   437.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Day18
ancova18<-aov(weight~Diet.x+base, data=day18_0)
summary(ancova18)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet.x       3  36690   12230   4.729 0.00623 **
## base         1   6229    6229   2.409 0.12818
## Residuals   42 108612    2586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Day20
ancova20<-aov(weight~Diet.x+base, data=day20_0)
summary(ancova20)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet.x       3  55881   18627   5.594 0.00261 **
## base         1   6672    6672   2.004 0.16447
## Residuals   41 136519    3330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value of Diet for 3 groups are lower than 0.05, we conclude that there is a significant difference between each group on Day10, Day18 and Day20.

2. Perform an appropriate repeated measures ANOVA, adjusting for baseline, to determine whether there is a significant difference in the mean weights of the four groups using the measurements on Days 10, 18, and 20. Do the analyses assuming compound symmetry and unstructured covariance structures and compare the results.

```
data<-rbind(day10_0, day18_0, day20_0)
colnames(data)<-c("weight", "base", "Time", "Chick", "Diet")
head(data)
```

```
##   weight base Time Chick Diet
## 1     93   42   10     1    1
## 2    103   40   10     2    1
## 3     99   43   10     3    1
## 4     87   42   10     4    1
## 5    106   41   10     5    1
## 6    124   41   10     6    1
```

```
#In order to assume compound symmetry and
#unstructured covariance structures, we check assumptions

#Check normality of residuals
data%>%group_by("Time", "Diet")%>%shapiro_test(weight)
```

```
## # A tibble: 1 x 5
##   `"Time"` `"Diet"` variable statistic        p
##   <chr>    <chr>    <chr>        <dbl>    <dbl>
## 1 Time     Diet     weight       0.956 0.000176
```

```
#p-value < 0.05, we assume normality of residuals

#Check homogeneity of variances
data%>%group_by(Time)%>%levene_test(weight~Diet)
```

```
## # A tibble: 3 x 5
##    Time   df1   df2 statistic     p
##   <dbl> <int> <int>     <dbl> <dbl>
## 1    10     3    45      1.39 0.259
```

```
## 2    18     3    43      1.06 0.376
## 3    20     3    42      1.11 0.355
```

```
#p-value>0.05, we assume homogeneity of variances

#Check outliers
data%>%group_by("Time", "Diet")%>%identify_outliers(weight)
```

```
## # A tibble: 1 x 9
##   `"Time"` `"Diet"` weight  base  Time Chick Diet  is.outlier is.extreme
##   <chr>    <chr>     <dbl> <dbl> <dbl> <ord> <fct> <lgl>      <lgl>
## 1 Time     Diet        361    41    20 35    3     TRUE       FALSE
```

```
#The outlier is not the extreme outlier.

res.aov<-aov(weight ~ base+Time*Diet+Error(Chick), data=data)
summary(res.aov)
```

```
##
## Error: Chick
##           Df Sum Sq Mean Sq F value Pr(>F)
## base       1  30068   30068   6.050 0.0181 *
## Time       1  25644   25644   5.160 0.0283 *
## Diet       3  62834   20945   4.214 0.0108 *
## Time:Diet  1    686     686   0.138 0.7121
## Residuals 42 208740    4970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Time       1 259282  259282 418.759  < 2e-16 ***
## Time:Diet  3  16422    5474   8.841 3.43e-05 ***
## Residuals 89  55106     619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(emmeans)
pairwise<-data%>%group_by(Time)%>%pairwise_t_test(weight~Diet,
                                        p.adjust.method = "bonferroni")
pairwise
```

```
## # A tibble: 18 x 10
##     Time .y.    group1 group2    n1    n2        p p.signif   p.adj p.adj.signif
##   * <dbl> <chr>  <chr>  <chr>  <int> <int>    <dbl> <chr>      <dbl> <chr>
##  1    10 weight 1      2         19    10 0.0626   ns       0.376   ns
##  2    10 weight 1      3         19    10 0.00474  **       0.0284  *
##  3    10 weight 2      3         10    10 0.358    ns       1       ns
##  4    10 weight 1      4         19    10 0.000186 ***      0.00112 **
##  5    10 weight 2      4         10    10 0.0653   ns       0.392   ns
##  6    10 weight 3      4         10    10 0.342    ns       1       ns
##  7    18 weight 1      2         17    10 0.17     ns       1       ns
##  8    18 weight 1      3         17    10 0.000816 ***      0.0049  **
##  9    18 weight 2      3         10    10 0.056    ns       0.336   ns
## 10    18 weight 1      4         17    10 0.0386   *        0.231   ns
## 11    18 weight 2      4         10    10 0.514    ns       1       ns
```

```
## 12    18 weight 3      4         10    10 0.198    ns       1       ns
## 13    20 weight 1      2         17    10 0.138    ns       0.828   ns
## 14    20 weight 1      3         17    10 0.000457 ***      0.00274 **
## 15    20 weight 2      3         10    10 0.0475   *        0.285   ns
## 16    20 weight 1      4         17     9 0.0117   *        0.07    ns
## 17    20 weight 2      4         10     9 0.298    ns       1       ns
## 18    20 weight 3      4         10     9 0.357    ns       1       ns
```

According to the result, we can conclude that: On Day 10: there is a significant difference in the mean weights of group 1 and group 3, group 1 and group 4,

On Day 18: there is a significant difference in the mean weights of group 1 and group 3, group 1 and group 4,

On Day 20: there is a significant difference in the mean weights of group 1 and group 3, group 2 and group 3, group 1 and group 4.