Advanced Data Analysis
Haoyang Chen | hc2812 | Assignment 5

1. Determine whether there is a significant difference in the mean weights of the six diet groups, using a one-way ANOVA

   a). Without adjusting for week 3 weight:
   p-value > 0.05, so there is no significant difference in the mean weights of the six diet groups

```
> summary(anov1)
            Df Sum Sq Mean Sq F value Pr(>F)
trt          1   14.7  14.672   2.366   0.13
Residuals   53  328.7   6.201
```

   b). Adjusting for Week 3 weight. Give the LS, and compare the results with (1a):

   p-value = 0.0137 < 0.05, there is significant difference in the mean weights of the six diet groups, the LS Means is 25.21021

```
            Df Sum Sq Mean Sq F value   Pr(>F)
covar        1  92.90   92.90  21.707 2.24e-05 ***
trt          1  27.89   27.89   6.518   0.0137 *
Residuals   52 222.55    4.28
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> lsmeans(diet.lm, 'covar')
 covar   lsmean        SE df lower.CL upper.CL
    29 25.21012 0.2789502 52 24.65037 25.76988

Confidence level used: 0.95
```

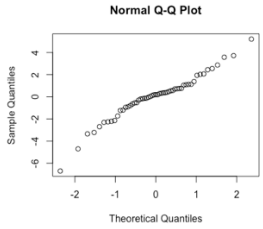
   c). Evaluate the appropriateness of performing inference based the adjusted means:

   p-value = 0.8130, inference about the marginal mean differences don’t need to be performed for each X=x. It is appropriate to perform inference based on the adjusted mean.

```
> summary(anov3)
            Df Sum Sq Mean Sq F value   Pr(>F)
covar        1  92.90   92.90  21.313 2.67e-05 ***
trt          1  27.89   27.89   6.399   0.0145 *
covar:trt    1   0.25    0.25   0.057   0.8130
Residuals   51 222.30    4.36
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

d). Check the validity of your assumptions:

(1). Normality:
Normality assumption is satisfied.
If it's not satisfied, we can use transformation, or robust procedure



(2). Homogeneity of variances:

```
> bartlett.test(diet.data$dmi, diet.data$trt)

        Bartlett test of homogeneity of variances

data:  diet.data$dmi and diet.data$trt
Bartlett's K-squared = 7.9927, df = 5, p-value = 0.1566
```

Homogeneity of variance assumption is satisfied.
If it's not satisfied, we can use transformation

(3). Parallelism:
From (c), we could know that parallelism is satisfied.
If not satisfied, inference about the marginal mean differences must be performed for each X = x

2. Comment on the use of the "average dmi during subsequent weeks" as a response variable

"Average dmi during subsequent weeks" depends on the number of subsequent weeks, so it's not a good response variable. Intuitively, "average dmi during subsequent weeks / the number of subsequent weeks" should be an appropriate response variable.

There is no significant difference in the mean weights of the six diet groups without adjusting for week 3 weight, but there is significant difference in the mean weights of the six diet groups with adjusting for week 3 weight.

```
> summary(anova4)
            Df Sum Sq Mean Sq F value Pr(>F)
trt          1  0.002 0.00163   0.012  0.915
Residuals   53  7.502 0.14154

> summary(anova5)
            Df Sum Sq Mean Sq F value   Pr(>F)
covar_avg    1  3.219   3.219  39.905 6.06e-08 ***
trt          1  0.090   0.090   1.121    0.295
Residuals   52  4.194   0.081
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Code:

```
diet.data <- read.table('diet.dat', header = TRUE)
diet.data <- diet.data[-c(18, 43),]


# without adjusting for week 3 weight
anov1 <- aov(dmi ~ trt, data = diet.data)
summary(anov1)

# adjusting for week 3 weight
diet.data$covar <- as.numeric(diet.data$covar)
anov2 <- aov(dmi ~ covar + trt, data = diet.data)
summary(anov2)

library(lsmeans)
diet.lm <- lm(dmi ~ covar + trt, data = diet.data)
summary(diet.lm)
lsmeans(diet.lm, 'covar')


# Test for parallelism
anov3 <- aov(dmi ~ covar*trt, data = diet.data)
summary(anov3)

# validation
qqnorm(diet.lm$residuals)
bartlett.test(diet.data$dmi, diet.data$trt)

# 2
diet.data$dmi_avg = diet.data$dmi/diet.data$weeks
diet.data$covar_avg = diet.data$covar/diet.data$weeks
anova4 = aov(dmi_avg~trt, data = diet.data)
summary(anova4)
anova5 = aov(dmi_avg~covar_avg + trt, data = diet.data)
summary(anova5)
```