Advanced Data Analysis

Haoyang Chen | hc2812 | Assignment 6

1. Do the analyses assuming compound symmetry, unstructured and AR(1) covariance structures and compare the results.

a). Compound Symmetry:

> anova(fit.cs)

Denom. DF: 220

numDF F-value p-value

(Intercept) 1 1125.1238 <.0001

Diet 3 7.3713 1e-04

Time 4 255.4254 <.0001

Diet:Time 12 4.0659 <.0001

b). Unconstructed

> anova(fit.un)

Denom. DF: 220

numDF F-value p-value

(Intercept) 1 18685.611 <.0001

Diet 3 9.377 <.0001

Time 4 137.297 <.0001

Diet:Time 12 5.218 <.0001

c). AR(1) Covariance Structure

> anova(fit.ar1)

Denom. DF: 220

numDF F-value p-value

(Intercept) 1 706.9114 <.0001

Diet 3 5.4191 0.0013

Time 4 213.6314 <.0001

Diet:Time 12 4.5165 <.0001

The p-values in the three covariance structures assumptions are < 0.05, thus there is a

significant difference in the mean weights of the four groups using the measurements on

Days 4, 8,12, 16 and 20.

d). Model Comparison:

> anova(fit.cs,fit.un)

Model df AIC BIC logLik Test L.Ratio p-value

fit.cs 1 22 2215.694 2290.354 -1085.8471

fit.un 2 35 1727.939 1846.716 -828.9696 1 vs 2 513.755 <.0001

> anova(fit.cs,fit.ar1)

Model df AIC BIC logLik

fit.cs 1 22 2215.694 2290.354 -1085.847

fit.ar1 2 22 2067.107 2141.766 -1011.553

Based on the result, the unconstructed covariance structure may be more suitable for this data as its AIC are lowest and Log-likelihood are highest.

1. In each case determine whether it might be appropriate to adjust for Birth Weight

a). Compound Symmetry:

> anova(fit.cs1)

Denom. DF: 219

numDF F-value p-value

(Intercept) 1 1136.2864 <.0001

Diet 3 7.4424 0.0001

Time 4 255.4010 <.0001

BirthWeight 1 1.5354 0.2166

Diet:Time 12 4.0547 <.0001

b). Unconstructed

> anova(fit.un1)

Denom. DF: 219

numDF F-value p-value

(Intercept) 1 19377.395 <.0001

Diet 3 9.852 <.0001

Time 4 137.174 <.0001

BirthWeight 1 3.164 0.0767

Diet:Time 12 5.218 <.0001

c). AR(1) Covariance Structure

> anova(fit.ar11)

Denom. DF: 219

numDF F-value p-value

(Intercept) 1 707.5728 <.0001

Diet 3 5.4240 0.0013

Time 4 213.7039 <.0001

BirthWeight 1 1.1872 0.2771

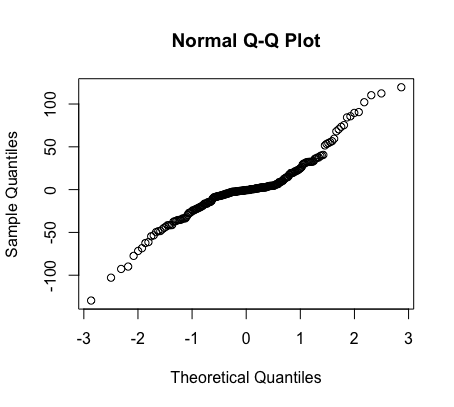
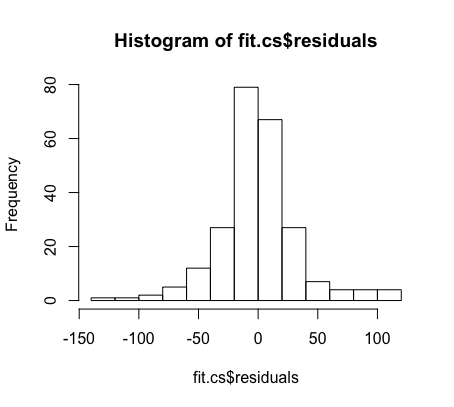
Diet:Time 12 4.5091 <.0001

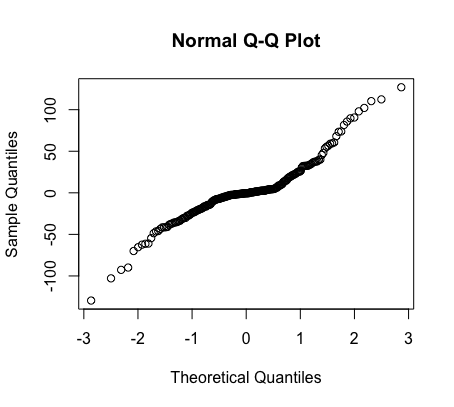
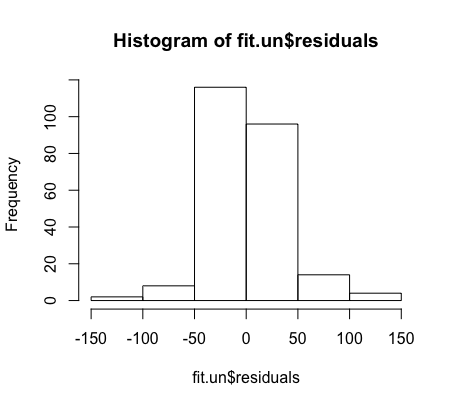
The p-values for the birthweight are greater than 0.05, which means that it is not significant to adjust for birth weight.

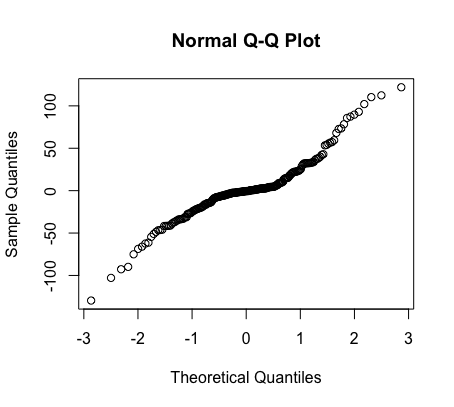
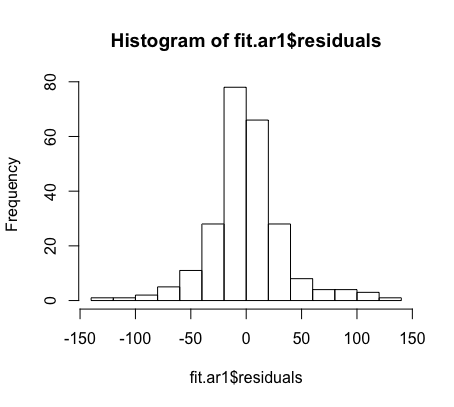
3.Check the validity of your assumptions

1). Normality

The normality assumption is not well satisfied for all three models







2). Homogeneity of Variance

Homogeneity of variance assumption is not satisfied

Bartlett test of homogeneity of variances

data: ChickWeight.filter$weight and ChickWeight.filter$Diet

Bartlett's K-squared = 10.472, df = 3, p-value = 0.01496

3). Parallelism

> summary(aov(ChickWeight.filter$weight ~ ChickWeight.filter$Diet\*ChickWeight.filter$Time))

Df Sum Sq Mean Sq

ChickWeight.filter$Diet 3 73462 24487

ChickWeight.filter$Time 4 667841 166960

ChickWeight.filter$Diet:ChickWeight.filter$Time 12 30033 2503

Residuals 220 272910 1240

F value Pr(>F)

ChickWeight.filter$Diet 19.740 2.28e-11 \*\*\*

ChickWeight.filter$Time 134.591 < 2e-16 \*\*\*

ChickWeight.filter$Diet:ChickWeight.filter$Time 2.018 0.0239 \*

Residuals

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Parallelism is not satisfied

R Code:

library(dplyr)

library(MASS)

data("ChickWeight")

ChickWeight$Diet <- factor(ChickWeight$Diet)

ChickWeight$Time <- factor(ChickWeight$Time)

ChickWeight$Chick <- factor(ChickWeight$Chick)

ChickWeight.filter <- filter(ChickWeight,Time==4 | Time==8 |Time==12 | Time==16 | Time==20)

# 1

library(nlme)

ChickWeight.repeat <- groupedData(weight~as.numeric(Diet)\*as.numeric(Time)|Chick,data=ChickWeight.filter)

fit.cs <- gls(weight ~ Diet \* Time, data=ChickWeight.repeat, corr=corCompSymm(,form=~1|Chick))

anova(fit.cs)

fit.un <- gls(weight~Diet\*Time,data=ChickWeight.repeat,corr=corSymm(form = ~1|Chick),weights = varIdent(form = ~1|Time))

anova(fit.un)

fit.ar1 <- gls(weight~Diet\*Time,data=ChickWeight.repeat,corr=corAR1(,form=~1|Chick))

anova(fit.ar1)

anova(fit.cs,fit.un)

anova(fit.cs,fit.ar1)

# 2

BirthWeight <- ChickWeight$weight

ChickWeight.birth <- cbind(ChickWeight,BirthWeight)

for (i in 1 : nrow(ChickWeight.birth)){

chick.index <- ChickWeight.birth$Chick[i]

ChickWeight.birth$BirthWeight[i] <-

ChickWeight.birth$weight[which(ChickWeight.birth$Chick == chick.index

& ChickWeight.birth$Time == 0)]

}

ChickWeight.birth\_filter <- filter(ChickWeight.birth,Time==4 | Time==8 |Time==12 | Time==16 | Time==20)

ChickWeight.birth\_repeated <- groupedData(weight~as.numeric(Diet)\*as.numeric(Time)|Chick,data=ChickWeight.birth\_filter)

fit.cs1 <- gls(weight~Diet\*Time+BirthWeight,data=ChickWeight.birth\_repeated,corr=corCompSymm(,form=~1|Chick))

anova(fit.cs1)

fit.un1 <- gls(weight~Diet\*Time+BirthWeight,data=ChickWeight.birth\_repeated,corr=corSymm(form = ~1|Chick),weights = varIdent(form = ~1|Time))

anova(fit.un1)

fit.ar11 <- gls(weight~Diet\*Time+BirthWeight,data=ChickWeight.birth\_repeated,corr=corAR1(,form=~1|Chick))

anova(fit.ar11)

# 3

qqnorm(fit.cs$residuals)

hist(fit.cs$residuals)

qqnorm(fit.un$residuals)

hist(fit.un$residuals)

qqnorm(fit.ar1$residuals)

hist(fit.ar1$residuals)

bartlett.test(x=ChickWeight.filter$weight,g=ChickWeight.filter$Diet)

summary(aov(ChickWeight.filter$weight ~ ChickWeight.filter$Diet\*ChickWeight.filter$Time))