HW2 BIOST540

Meixin Zhang

4/21/2021

## Part A

rm(list=ls())  
library(reshape2)

## Warning: package 'reshape2' was built under R version 3.6.2

library(ggplot2)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(nlme)

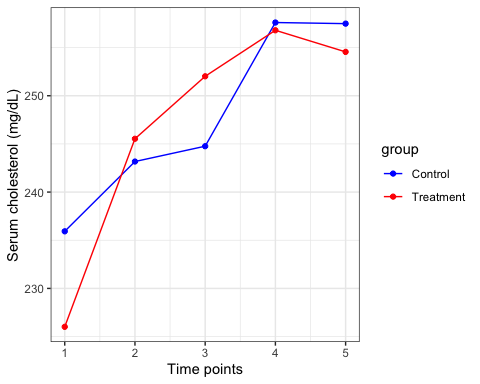
##   
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':  
##   
## collapse

cholesterol <- read.csv("~/Documents/MS-2nd-year/Spring/BIOST540/HW/cholesterol.csv")

## Question 1

cholesterol<-cholesterol[,c("group","id","y1","y2","y3","y4","y5")]  
choles\_long <- melt(cholesterol, id=c("id","group"))  
names(choles\_long)<-c("id","group","time","value")  
choles\_long$time <- as.numeric(gsub("y","",choles\_long$time))  
choles\_long$group[choles\_long$group==1]<-"Treatment"  
choles\_long$group[choles\_long$group==2]<-"Control"  
choles\_means <- aggregate(value ~ time + group, data = choles\_long, FUN=mean)  
p <- ggplot(data = choles\_means, aes(x = time, y = value, group = group,   
 col = group))  
p + geom\_line() + scale\_color\_manual(values=c('blue','red')) +   
 geom\_point() + theme\_bw() + ylab("Serum cholesterol (mg/dL)") + xlab("Time points")



## summary statistics

by(cholesterol[,-c(1,2)], INDICES = cholesterol$group, FUN=summary)

## cholesterol$group: 1  
## y1 y2 y3 y4   
## Min. :144.0 Min. :177.0 Min. :167.0 Min. :194.0   
## 1st Qu.:200.2 1st Qu.:214.0 1st Qu.:223.5 1st Qu.:232.8   
## Median :222.0 Median :248.5 Median :260.0 Median :254.0   
## Mean :226.0 Mean :245.5 Mean :252.0 Mean :256.8   
## 3rd Qu.:249.5 3rd Qu.:271.8 3rd Qu.:277.5 3rd Qu.:275.0   
## Max. :313.0 Max. :334.0 Max. :316.0 Max. :334.0   
## NA's :7 NA's :18   
## y5   
## Min. :172.0   
## 1st Qu.:214.0   
## Median :251.0   
## Mean :254.6   
## 3rd Qu.:283.0   
## Max. :397.0   
## NA's :24   
## --------------------------------------------------------   
## cholesterol$group: 2  
## y1 y2 y3 y4   
## Min. :141.0 Min. :142.0 Min. :157.0 Min. :162.0   
## 1st Qu.:197.0 1st Qu.:214.0 1st Qu.:215.2 1st Qu.:223.0   
## Median :230.0 Median :242.0 Median :238.0 Median :259.0   
## Mean :235.9 Mean :243.2 Mean :244.8 Mean :257.6   
## 3rd Qu.:262.0 3rd Qu.:274.0 3rd Qu.:276.0 3rd Qu.:286.5   
## Max. :418.0 Max. :371.0 Max. :363.0 Max. :384.0   
## NA's :3 NA's :6   
## y5   
## Min. :169.0   
## 1st Qu.:227.5   
## Median :248.0   
## Mean :257.5   
## 3rd Qu.:279.0   
## Max. :387.0   
## NA's :10

## Estimated correlation matrix after subtracting out means by group and month

choles\_mean <- cholesterol %>% group\_by(group) %>%   
 summarise(y0mean = mean(y1), y6mean=mean(y2),   
 y12mean=mean(y3,na.rm = TRUE), y20mean=mean(y4,na.rm = TRUE),  
 y24mean=mean(y5,na.rm = TRUE))   
cholesterol <- merge(cholesterol, choles\_mean, by="group")  
cholesterol$y0.resid <- cholesterol$y1 - cholesterol$y0mean  
cholesterol$y6.resid <- cholesterol$y2 - cholesterol$y6mean  
cholesterol$y12.resid <- cholesterol$y3 - cholesterol$y12mean  
cholesterol$y20.resid <- cholesterol$y4 - cholesterol$y20mean  
cholesterol$y24.resid <- cholesterol$y5 - cholesterol$y24mean  
choles\_complete<-cholesterol[rowSums(is.na(cholesterol))==0,]  
cor(choles\_complete[,c("y0.resid", "y6.resid", "y12.resid", "y20.resid","y24.resid")])

## y0.resid y6.resid y12.resid y20.resid y24.resid  
## y0.resid 1.0000000 0.7634717 0.7479556 0.7547128 0.6055315  
## y6.resid 0.7634717 1.0000000 0.8066409 0.8132653 0.6946468  
## y12.resid 0.7479556 0.8066409 1.0000000 0.7373536 0.7035717  
## y20.resid 0.7547128 0.8132653 0.7373536 1.0000000 0.6446018  
## y24.resid 0.6055315 0.6946468 0.7035717 0.6446018 1.0000000

## Estimated correlation matrix by treatment groups

by(choles\_complete[,c("y0.resid", "y6.resid", "y12.resid", "y20.resid", "y24.resid")],   
 INDICES = choles\_complete$group, FUN=cor)

## choles\_complete$group: 1  
## y0.resid y6.resid y12.resid y20.resid y24.resid  
## y0.resid 1.0000000 0.7097680 0.6590338 0.6345956 0.4515519  
## y6.resid 0.7097680 1.0000000 0.6794303 0.7294584 0.5739744  
## y12.resid 0.6590338 0.6794303 1.0000000 0.5218361 0.6363163  
## y20.resid 0.6345956 0.7294584 0.5218361 1.0000000 0.5070192  
## y24.resid 0.4515519 0.5739744 0.6363163 0.5070192 1.0000000  
## --------------------------------------------------------   
## choles\_complete$group: 2  
## y0.resid y6.resid y12.resid y20.resid y24.resid  
## y0.resid 1.0000000 0.8041079 0.8167669 0.8390164 0.7612846  
## y6.resid 0.8041079 1.0000000 0.9046082 0.8824122 0.8191013  
## y12.resid 0.8167669 0.9046082 1.0000000 0.8884498 0.7776534  
## y20.resid 0.8390164 0.8824122 0.8884498 1.0000000 0.7892396  
## y24.resid 0.7612846 0.8191013 0.7776534 0.7892396 1.0000000

## Question 3

m.reml<-gls(value~group\*as.factor(time),  
 data=choles\_long[rowSums(is.na(choles\_long))==0,],  
 method="REML",  
 correlation=corSymm(form = ~time | id),  
 weights=varIdent(form= ~1 | time))  
summary(m.reml)

## Generalized least squares fit by REML  
## Model: value ~ group \* as.factor(time)   
## Data: choles\_long[rowSums(is.na(choles\_long)) == 0, ]   
## AIC BIC logLik  
## 4315.08 4417.078 -2132.54  
##   
## Correlation Structure: General  
## Formula: ~time | id   
## Parameter estimate(s):  
## Correlation:   
## 1 2 3 4   
## 2 0.770   
## 3 0.732 0.775   
## 4 0.737 0.796 0.725   
## 5 0.589 0.669 0.680 0.626  
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | time   
## Parameter estimates:  
## 1 2 3 4 5   
## 1.0000000 0.9322229 0.8798471 0.8962106 1.0303536   
##   
## Coefficients:  
## Value Std.Error t-value p-value  
## (Intercept) 235.92683 7.302782 32.30643 0.0000  
## groupTreatment -9.91070 9.412632 -1.05291 0.2930  
## as.factor(time)2 7.24390 4.805468 1.50743 0.1324  
## as.factor(time)3 8.84832 5.207268 1.69922 0.0900  
## as.factor(time)4 23.10279 5.297447 4.36112 0.0000  
## as.factor(time)5 21.12377 7.369940 2.86621 0.0044  
## groupTreatment:as.factor(time)2 12.27223 6.193818 1.98137 0.0482  
## groupTreatment:as.factor(time)3 16.41754 6.743418 2.43460 0.0153  
## groupTreatment:as.factor(time)4 4.97699 6.982010 0.71283 0.4763  
## groupTreatment:as.factor(time)5 6.90311 9.791182 0.70503 0.4812  
##   
## Correlation:   
## (Intr) grpTrt as.()2 as.()3 as.()4 as.()5  
## groupTreatment -0.776   
## as.factor(time)2 -0.429 0.332   
## as.factor(time)3 -0.500 0.388 0.584   
## as.factor(time)4 -0.468 0.363 0.601 0.525   
## as.factor(time)5 -0.390 0.303 0.480 0.522 0.437   
## groupTreatment:as.factor(time)2 0.332 -0.429 -0.776 -0.453 -0.466 -0.372  
## groupTreatment:as.factor(time)3 0.386 -0.497 -0.451 -0.772 -0.405 -0.403  
## groupTreatment:as.factor(time)4 0.355 -0.458 -0.456 -0.398 -0.759 -0.332  
## groupTreatment:as.factor(time)5 0.293 -0.378 -0.361 -0.393 -0.329 -0.753  
## gT:.()2 gT:.()3 gT:.()4  
## groupTreatment   
## as.factor(time)2   
## as.factor(time)3   
## as.factor(time)4   
## as.factor(time)5   
## groupTreatment:as.factor(time)2   
## groupTreatment:as.factor(time)3 0.582   
## groupTreatment:as.factor(time)4 0.587 0.512   
## groupTreatment:as.factor(time)5 0.465 0.504 0.419   
##   
## Standardized residuals:  
## Min Q1 Med Q3 Max   
## -2.32089174 -0.68934564 -0.02685699 0.61014657 3.89372888   
##   
## Residual standard error: 46.76062   
## Degrees of freedom: 447 total; 437 residual

anova(m.reml)

## Denom. DF: 437   
## numDF F-value p-value  
## (Intercept) 1 4280.329 <.0001  
## group 1 0.006 0.9369  
## as.factor(time) 4 16.471 <.0001  
## group:as.factor(time) 4 1.965 0.0988

## Question 4

choles\_long$month[choles\_long$time==1]<-0  
choles\_long$month[choles\_long$time==2]<-6  
choles\_long$month[choles\_long$time==3]<-12  
choles\_long$month[choles\_long$time==4]<-20  
choles\_long$month[choles\_long$time==5]<-24  
m.reml2<-gls(value~group\*month,  
 data=choles\_long[rowSums(is.na(choles\_long))==0,],  
 method="REML",  
 correlation=corSymm(form = ~time | id),  
 weights=varIdent(form= ~1 | month))  
summary(m.reml2)

## Generalized least squares fit by REML  
## Model: value ~ group \* month   
## Data: choles\_long[rowSums(is.na(choles\_long)) == 0, ]   
## AIC BIC logLik  
## 4360.762 4438.54 -2161.381  
##   
## Correlation Structure: General  
## Formula: ~time | id   
## Parameter estimate(s):  
## Correlation:   
## 1 2 3 4   
## 2 0.753   
## 3 0.717 0.778   
## 4 0.736 0.790 0.718   
## 5 0.593 0.654 0.678 0.632  
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | month   
## Parameter estimates:  
## 0 6 12 20 24   
## 1.0000000 0.9262840 0.8751178 0.8863003 1.0245799   
##   
## Coefficients:  
## Value Std.Error t-value p-value  
## (Intercept) 235.28055 6.779242 34.70603 0.0000  
## groupTreatment -0.86560 8.750192 -0.09892 0.9212  
## month 1.01972 0.218920 4.65794 0.0000  
## groupTreatment:month 0.09761 0.290038 0.33654 0.7366  
##   
## Correlation:   
## (Intr) grpTrt month   
## groupTreatment -0.775   
## month -0.461 0.357   
## groupTreatment:month 0.348 -0.459 -0.755  
##   
## Standardized residuals:  
## Min Q1 Med Q3 Max   
## -2.2757429 -0.7103864 -0.0483643 0.6376086 3.8749918   
##   
## Residual standard error: 47.15351   
## Degrees of freedom: 447 total; 443 residual

## Part B

data(Orthodont)  
Orthodont$id<-cumsum(!duplicated(Orthodont$Subject))

## Question 2

## OLS

lm<-lm(distance ~ as.factor(age)\*Sex,   
 data = Orthodont)  
summary(lm)

##   
## Call:  
## lm(formula = distance ~ as.factor(age) \* Sex, data = Orthodont)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.8750 -1.3984 -0.1818 1.4091 5.2813   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 22.8750 0.5734 39.894 < 2e-16 \*\*\*  
## as.factor(age)10 0.9375 0.8109 1.156 0.250384   
## as.factor(age)12 2.8438 0.8109 3.507 0.000681 \*\*\*  
## as.factor(age)14 4.5938 0.8109 5.665 1.42e-07 \*\*\*  
## SexFemale -1.6932 0.8983 -1.885 0.062359 .   
## as.factor(age)10:SexFemale 0.1080 1.2704 0.085 0.932451   
## as.factor(age)12:SexFemale -0.9347 1.2704 -0.736 0.463634   
## as.factor(age)14:SexFemale -1.6847 1.2704 -1.326 0.187843   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.294 on 100 degrees of freedom  
## Multiple R-squared: 0.4268, Adjusted R-squared: 0.3867   
## F-statistic: 10.64 on 7 and 100 DF, p-value: 6.075e-10

## GLS, unstructured/symmetric correlation matrix, heteroscedasticity, REML

Orthodont$time<-ifelse(Orthodont$age==8,1,  
 ifelse(Orthodont$age==10,2,  
 ifelse(Orthodont$age==12,3,4)))  
reml\_2<-gls(distance ~ Sex\*as.factor(age),  
 data = Orthodont,  
 method = "REML",  
 correlation = corSymm(form = ~time | Subject),  
 weights = varIdent(form = ~1 | as.factor(age)))  
summary(reml\_2)

## Generalized least squares fit by REML  
## Model: distance ~ Sex \* as.factor(age)   
## Data: Orthodont   
## AIC BIC logLik  
## 450.0348 496.9279 -207.0174  
##   
## Correlation Structure: General  
## Formula: ~time | Subject   
## Parameter estimate(s):  
## Correlation:   
## 1 2 3   
## 2 0.571   
## 3 0.661 0.563   
## 4 0.522 0.726 0.728  
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | as.factor(age)   
## Parameter estimates:  
## 8 10 12 14   
## 1.0000000 0.8790601 1.0918305 0.9595050   
##   
## Coefficients:  
## Value Std.Error t-value p-value  
## (Intercept) 22.875000 0.5817783 39.31910 0.0000  
## SexFemale -1.693182 0.9114715 -1.85764 0.0662  
## as.factor(age)10 0.937500 0.5103055 1.83713 0.0692  
## as.factor(age)12 2.843750 0.5031617 5.65176 0.0000  
## as.factor(age)14 4.593750 0.5579390 8.23343 0.0000  
## SexFemale:as.factor(age)10 0.107955 0.7994951 0.13503 0.8929  
## SexFemale:as.factor(age)12 -0.934659 0.7883028 -1.18566 0.2386  
## SexFemale:as.factor(age)14 -1.684659 0.8741224 -1.92726 0.0568  
##   
## Correlation:   
## (Intr) SexFml a.()10 a.()12 a.()14 SF:.()10  
## SexFemale -0.638   
## as.factor(age)10 -0.568 0.363   
## as.factor(age)12 -0.321 0.205 0.418   
## as.factor(age)14 -0.521 0.332 0.726 0.651   
## SexFemale:as.factor(age)10 0.363 -0.568 -0.638 -0.267 -0.463   
## SexFemale:as.factor(age)12 0.205 -0.321 -0.267 -0.638 -0.416 0.418   
## SexFemale:as.factor(age)14 0.332 -0.521 -0.463 -0.416 -0.638 0.726   
## SF:.()12  
## SexFemale   
## as.factor(age)10   
## as.factor(age)12   
## as.factor(age)14   
## SexFemale:as.factor(age)10   
## SexFemale:as.factor(age)12   
## SexFemale:as.factor(age)14 0.651   
##   
## Standardized residuals:  
## Min Q1 Med Q3 Max   
## -2.52458690 -0.64159831 -0.07813035 0.59313659 2.07856659   
##   
## Residual standard error: 2.327113   
## Degrees of freedom: 108 total; 100 residual

## GLS, exchangeable/compound symmetric correlation matrix, homoscedasticity, REML

reml\_3<-gls(distance ~ Sex\*as.factor(age),   
 data=Orthodont,   
 method="REML",  
 correlation=corCompSymm(form = ~time | id))  
summary(reml\_3)

## Generalized least squares fit by REML  
## Model: distance ~ Sex \* as.factor(age)   
## Data: Orthodont   
## AIC BIC logLik  
## 443.4085 469.4602 -211.7043  
##   
## Correlation Structure: Compound symmetry  
## Formula: ~time | id   
## Parameter estimate(s):  
## Rho   
## 0.6245472   
##   
## Coefficients:  
## Value Std.Error t-value p-value  
## (Intercept) 22.875000 0.5733901 39.89430 0.0000  
## SexFemale -1.693182 0.8983297 -1.88481 0.0624  
## as.factor(age)10 0.937500 0.4968701 1.88681 0.0621  
## as.factor(age)12 2.843750 0.4968701 5.72333 0.0000  
## as.factor(age)14 4.593750 0.4968701 9.24537 0.0000  
## SexFemale:as.factor(age)10 0.107955 0.7784458 0.13868 0.8900  
## SexFemale:as.factor(age)12 -0.934659 0.7784458 -1.20067 0.2327  
## SexFemale:as.factor(age)14 -1.684659 0.7784458 -2.16413 0.0328  
##   
## Correlation:   
## (Intr) SexFml a.()10 a.()12 a.()14 SF:.()10  
## SexFemale -0.638   
## as.factor(age)10 -0.433 0.277   
## as.factor(age)12 -0.433 0.277 0.500   
## as.factor(age)14 -0.433 0.277 0.500 0.500   
## SexFemale:as.factor(age)10 0.277 -0.433 -0.638 -0.319 -0.319   
## SexFemale:as.factor(age)12 0.277 -0.433 -0.319 -0.638 -0.319 0.500   
## SexFemale:as.factor(age)14 0.277 -0.433 -0.319 -0.319 -0.638 0.500   
## SF:.()12  
## SexFemale   
## as.factor(age)10   
## as.factor(age)12   
## as.factor(age)14   
## SexFemale:as.factor(age)10   
## SexFemale:as.factor(age)12   
## SexFemale:as.factor(age)14 0.500   
##   
## Standardized residuals:  
## Min Q1 Med Q3 Max   
## -2.56151951 -0.60972339 -0.07927333 0.61436832 2.30264254   
##   
## Residual standard error: 2.293561   
## Degrees of freedom: 108 total; 100 residual

## LMM, random intercepts, REML

lmm1 <- lme( distance ~ Sex\*as.factor(age),   
 method = "REML", data = Orthodont,   
 random = reStruct( ~ 1 | id, pdClass="pdDiag", REML=F))  
summary(lmm1)

## Linear mixed-effects model fit by REML  
## Data: Orthodont   
## AIC BIC logLik  
## 443.4085 469.4602 -211.7043  
##   
## Random effects:  
## Formula: ~1 | id  
## (Intercept) Residual  
## StdDev: 1.812564 1.40536  
##   
## Fixed effects: distance ~ Sex \* as.factor(age)   
## Value Std.Error DF t-value p-value  
## (Intercept) 22.875000 0.5733905 75 39.89428 0.0000  
## SexFemale -1.693182 0.8983302 25 -1.88481 0.0711  
## as.factor(age)10 0.937500 0.4968699 75 1.88681 0.0631  
## as.factor(age)12 2.843750 0.4968699 75 5.72333 0.0000  
## as.factor(age)14 4.593750 0.4968699 75 9.24538 0.0000  
## SexFemale:as.factor(age)10 0.107955 0.7784456 75 0.13868 0.8901  
## SexFemale:as.factor(age)12 -0.934659 0.7784456 75 -1.20067 0.2337  
## SexFemale:as.factor(age)14 -1.684659 0.7784456 75 -2.16413 0.0336  
## Correlation:   
## (Intr) SexFml a.()10 a.()12 a.()14 SF:.()10  
## SexFemale -0.638   
## as.factor(age)10 -0.433 0.277   
## as.factor(age)12 -0.433 0.277 0.500   
## as.factor(age)14 -0.433 0.277 0.500 0.500   
## SexFemale:as.factor(age)10 0.277 -0.433 -0.638 -0.319 -0.319   
## SexFemale:as.factor(age)12 0.277 -0.433 -0.319 -0.638 -0.319 0.500   
## SexFemale:as.factor(age)14 0.277 -0.433 -0.319 -0.319 -0.638 0.500   
## SF:.()12  
## SexFemale   
## as.factor(age)10   
## as.factor(age)12   
## as.factor(age)14   
## SexFemale:as.factor(age)10   
## SexFemale:as.factor(age)12   
## SexFemale:as.factor(age)14 0.500   
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -3.73580847 -0.47606875 0.04687496 0.46911063 3.66127764   
##   
## Number of Observations: 108  
## Number of Groups: 27

## LMM, random intercepts + slopes (correlated), REML

lmm2 <- lme( distance ~ Sex\*as.factor(age),   
 method = "REML", data = Orthodont,   
 random = reStruct( ~ 1 + age | id, pdClass="pdSymm", REML=F))  
summary(lmm2)

## Linear mixed-effects model fit by REML  
## Data: Orthodont   
## AIC BIC logLik  
## 446.4494 477.7114 -211.2247  
##   
## Random effects:  
## Formula: ~1 + age | id  
## Structure: General positive-definite  
## StdDev Corr   
## (Intercept) 2.3215279 (Intr)  
## age 0.1713889 -0.641  
## Residual 1.3338704   
##   
## Fixed effects: distance ~ Sex \* as.factor(age)   
## Value Std.Error DF t-value p-value  
## (Intercept) 22.875000 0.5572853 75 41.04720 0.0000  
## SexFemale -1.693182 0.8730983 25 -1.93928 0.0638  
## as.factor(age)10 0.937500 0.4793170 75 1.95591 0.0542  
## as.factor(age)12 2.843750 0.5017723 75 5.66741 0.0000  
## as.factor(age)14 4.593750 0.5371155 75 8.55263 0.0000  
## SexFemale:as.factor(age)10 0.107955 0.7509455 75 0.14376 0.8861  
## SexFemale:as.factor(age)12 -0.934659 0.7861261 75 -1.18894 0.2382  
## SexFemale:as.factor(age)14 -1.684659 0.8414983 75 -2.00198 0.0489  
## Correlation:   
## (Intr) SexFml a.()10 a.()12 a.()14 SF:.()10  
## SexFemale -0.638   
## as.factor(age)10 -0.426 0.272   
## as.factor(age)12 -0.416 0.265 0.523   
## as.factor(age)14 -0.397 0.253 0.518 0.576   
## SexFemale:as.factor(age)10 0.272 -0.426 -0.638 -0.334 -0.330   
## SexFemale:as.factor(age)12 0.265 -0.416 -0.334 -0.638 -0.368 0.523   
## SexFemale:as.factor(age)14 0.253 -0.397 -0.330 -0.368 -0.638 0.518   
## SF:.()12  
## SexFemale   
## as.factor(age)10   
## as.factor(age)12   
## as.factor(age)14   
## SexFemale:as.factor(age)10   
## SexFemale:as.factor(age)12   
## SexFemale:as.factor(age)14 0.576   
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -3.38162828 -0.41119936 0.04314813 0.40300524 3.81039469   
##   
## Number of Observations: 108  
## Number of Groups: 27