STAT 645 HW 7

Nitish Neelagiri

November 13, 2015

```
setwd("~/")
exerciseData <- read.table("exercise.txt", na.strings = ".", header=FALSE)
names(exerciseData) <- c("ID","Program","0","2","4","6","8","10","12")
exDataNARM <- exerciseData[complete.cases(exerciseData),]
exDataNARM_long_q1 <- reshape(exDataNARM, idvar = "ID", varying = list(3:9),direction = "long")
names(exDataNARM_long_q1) <- c("ID", "Treatment", "Time", "Strength")
exDataNARM_long_q1$Time <- factor(c("0","2","4","6","8","10","12"))
attach(exDataNARM_long_q1)
fit <- lm(Strength~Time+Treatment, data=exDataNARM_long_q1)
summary(fit)
```

```
##
## Call:
## lm(formula = Strength ~ Time + Treatment, data = exDataNARM_long_q1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7984 -2.6679  0.2451  2.2451  6.0770
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 79.61280    1.08633  73.286   <2e-16 ***
## Time10       0.17391    0.98672   0.176   0.8603
## Time12      -0.21739    0.98672  -0.220   0.8259
## Time2       -0.13043    0.98672  -0.132   0.8950
## Time4       -0.04348    0.98672  -0.044   0.9649
## Time6        0.26087    0.98672   0.264   0.7918
## Time8       -0.08696    0.98672  -0.088   0.9299
## Treatment    1.13626    0.53197   2.136   0.0343 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.346 on 153 degrees of freedom
## Multiple R-squared:  0.03112,    Adjusted R-squared:  -0.0132
## F-statistic: 0.7022 on 7 and 153 DF,  p-value: 0.6702
```

```
ee_fit <- residuals(fit)
ee_fit_matrix <- matrix(ee_fit, nrow=23, ncol=7)
ee_fit_matrix
```

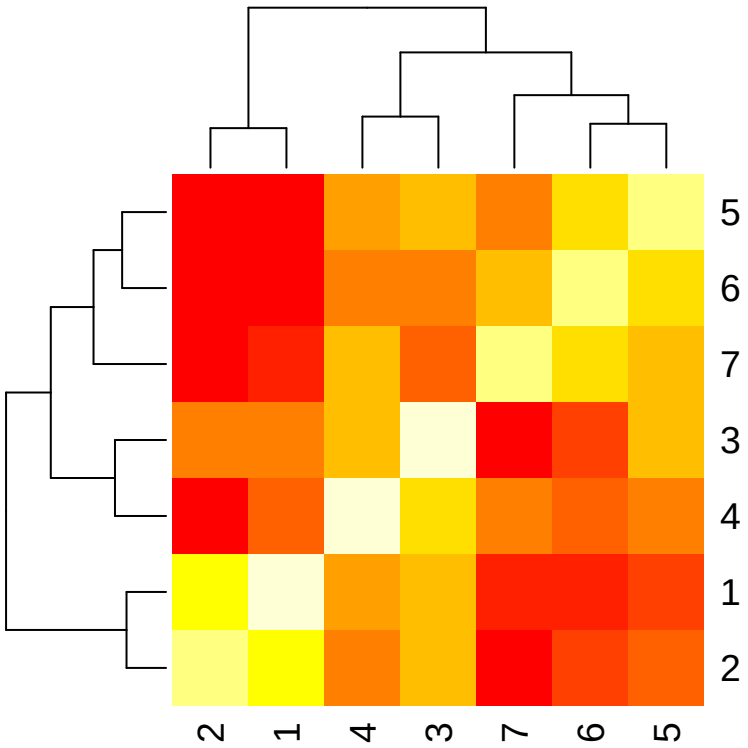
```
##             [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
##  [1,]  2.2509317  2.2944099  4.3378882  4.4683230  5.3813665  5.9900621
##  [2,]  0.3813665  1.9900621  1.0770186  1.2509317  2.2944099  2.3378882
##  [3,]  0.2944099  0.3378882  0.4683230  1.3813665  0.9900621  2.0770186
##  [4,] -5.0099379 -4.9229814 -4.7490683 -4.7055901 -4.6621118 -4.5316770
##  [5,]  0.3378882  3.4683230  2.3813665  1.9900621  4.0770186  4.2509317
```

```
##  [6,] -3.9229814 -2.7490683 -1.7055901 -1.6621118  0.4683230  1.3813665
##  [7,] -6.5316770 -5.6186335 -3.0099379 -2.9229814 -1.7490683 -2.7055901
##  [8,] -4.7490683 -3.7055901 -3.6621118 -3.5316770 -3.6186335 -5.0099379
##  [9,]  3.3813665  2.9900621  5.0770186  4.2509317  5.2944099  5.3378882
## [10,] -1.7055901 -0.6621118 -1.5316770 -0.6186335 -1.0099379  1.0770186
## [11,]  1.8537984  2.9407549  2.1146679  1.1581462  1.2016245  1.3320592
## [12,] -7.7983755 -6.6679408 -6.7548973 -4.1462016 -7.0592451 -5.8853321
## [13,]  3.9407549  5.1146679  5.1581462  5.2016245  5.3320592  5.2451027
## [14,]  0.3320592  1.2451027  1.8537984  2.9407549  2.1146679  3.1581462
## [15,] -2.8853321 -1.8418538 -2.7983755 -2.6679408 -1.7548973 -3.1462016
## [16,] -0.7548973 -1.1462016 -1.0592451  0.1146679  0.1581462  1.2016245
## [17,] -0.8418538 -0.7983755  0.3320592  2.2451027  0.8537984 -0.0592451
## [18,] -4.1462016 -4.0592451 -2.8853321 -2.8418538 -3.7983755 -2.6679408
## [19,]  1.2016245  0.3320592  0.2451027  1.8537984  1.9407549  1.1146679
## [20,] -2.0592451  0.1146679  0.1581462  0.2016245 -0.6679408 -0.7548973
## [21,]  3.3320592  4.2451027  4.8537984  3.9407549  4.1146679  4.1581462
## [22,] -4.8853321 -3.8418538 -1.7983755 -0.6679408  0.2451027 -0.1462016
## [23,] -1.7548973 -1.1462016 -2.0592451 -0.8853321 -0.8418538  0.2016245
##               [,7]
##  [1,]  6.077018634
##  [2,]  1.468322981
##  [3,]  0.250931677
##  [4,] -5.618633540
##  [5,]  4.294409938
##  [6,] -0.009937888
##  [7,] -2.662111801
##  [8,] -4.922981366
##  [9,]  5.468322981
## [10,]  1.250931677
## [11,]  2.245102723
## [12,] -5.841853798
## [13,]  3.853798376
## [14,]  4.201624462
## [15,] -2.059245103
## [16,]  1.332059245
## [17,]  3.114667941
## [18,] -2.754897277
## [19,]  2.158146202
## [20,] -1.146201624
## [21,]  4.201624462
## [22,] -0.059245103
## [23,]  1.332059245
```

```
cor(ee_fit_matrix)
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 1.0000000 0.9640790 0.9414688 0.9283516 0.8935351 0.8838908 0.8795459
## [2,] 0.9640790 1.0000000 0.9437685 0.9083287 0.8956858 0.8833133 0.8634304
## [3,] 0.9414688 0.9437685 1.0000000 0.9623066 0.9593483 0.9318536 0.9103415
## [4,] 0.9283516 0.9083287 0.9623066 1.0000000 0.9425692 0.9306989 0.9399982
## [5,] 0.8935351 0.8956858 0.9593483 0.9425692 1.0000000 0.9630699 0.9400965
## [6,] 0.8838908 0.8833133 0.9318536 0.9306989 0.9630699 1.0000000 0.9518939
## [7,] 0.8795459 0.8634304 0.9103415 0.9399982 0.9400965 0.9518939 1.0000000
```

```
heatmap(cor(ee_fit_matrix))
```

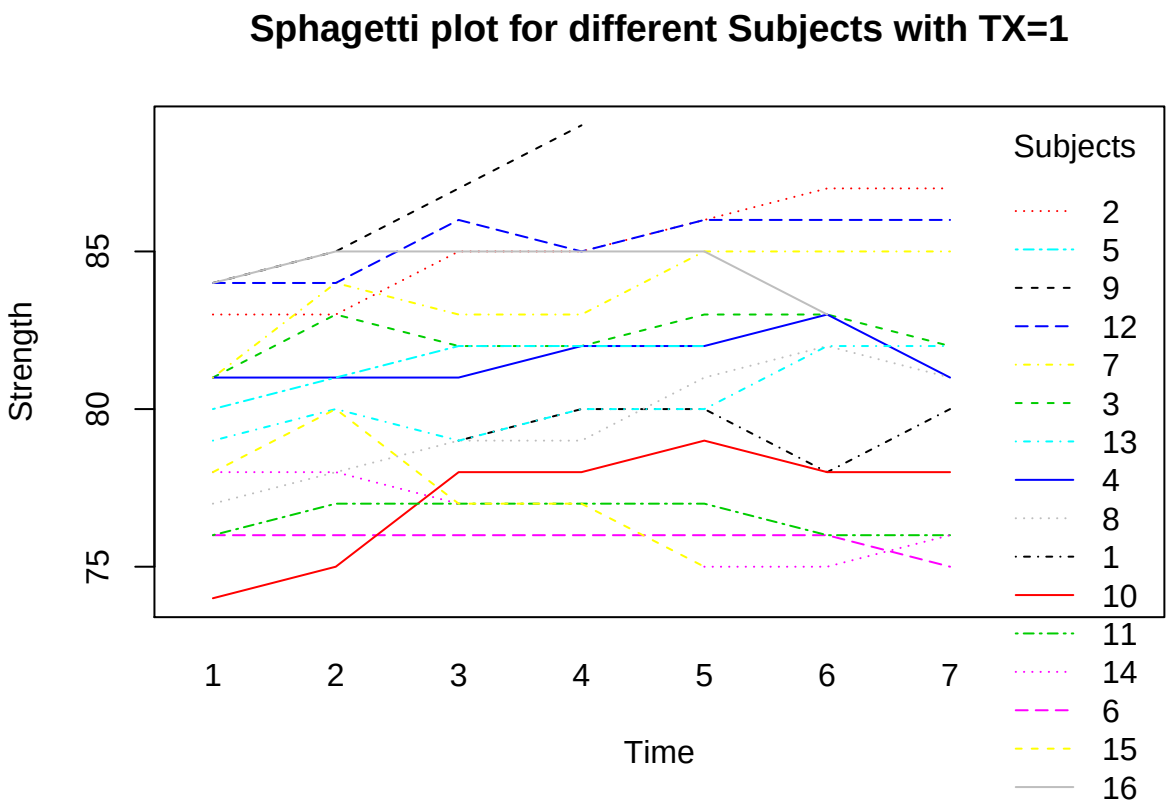


(b)

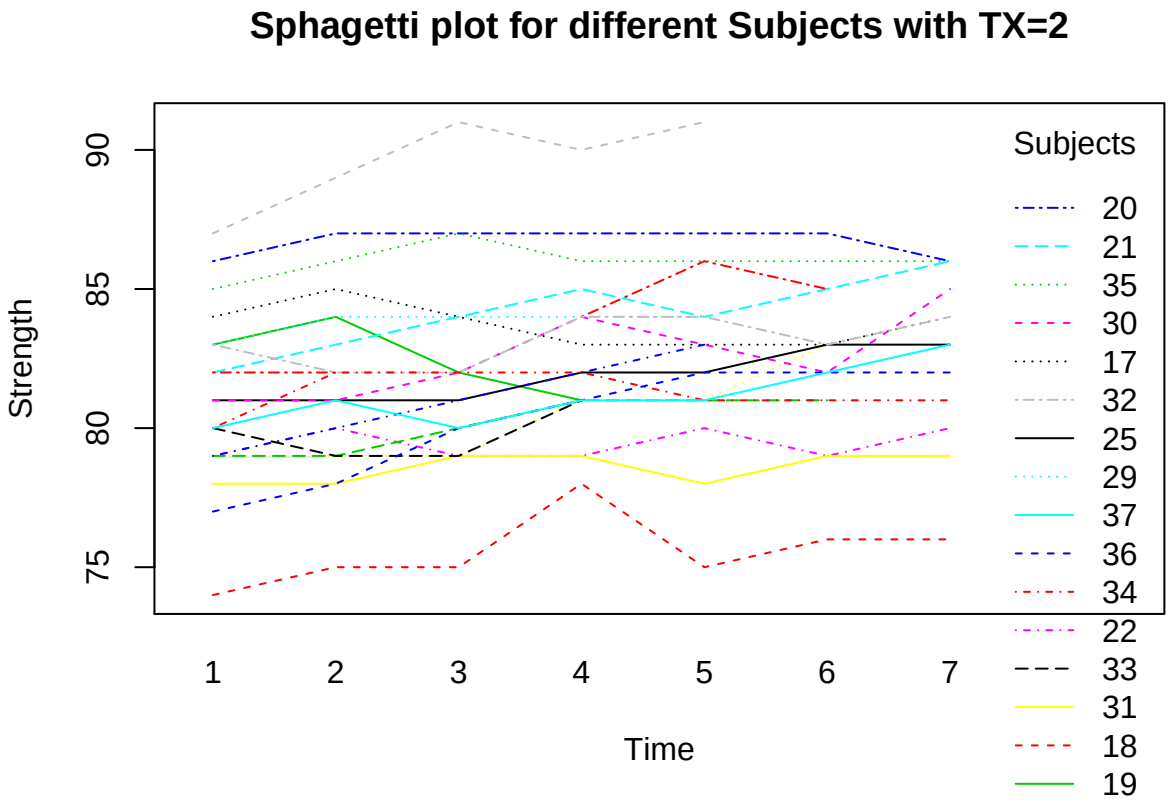
```
exDataNARM_long_q2 <- reshape(exerciseData, idvar = "ID", varying = list(3:9),direction = "long")
names(exDataNARM_long_q2) <- c("ID", "Treatment", "Time", "Strength")
attach(exDataNARM_long_q2)
```

```
## The following objects are masked from exDataNARM_long_q1:
##
##     ID, Strength, Time, Treatment
```

```
interaction.plot(Time[Treatment==1], ID[Treatment==1], Strength[Treatment==1], xlab = "Time", ylab = "S
```




```
interaction.plot(Time[Treatment==2], ID[Treatment==2], Strength[Treatment==2], xlab = "Time", ylab = "S
```



(c)

```r
library(gee)
dta_new <- NULL
for(i in 1:nrow(exDataNARM_long_q2))
  dta_new <- rbind(dta_new, exDataNARM_long_q2[exDataNARM_long_q2$ID == i, ])
fit_gee <- gee(Strength ~ Time + Treatment, id = ID, corstr = "AR-M", data=dta_new)
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
## running glm to get initial regression estimate

## (Intercept)        Time   Treatment
##  78.5447397   0.2196822   1.3551741
```

```r
summary(fit_gee)
```

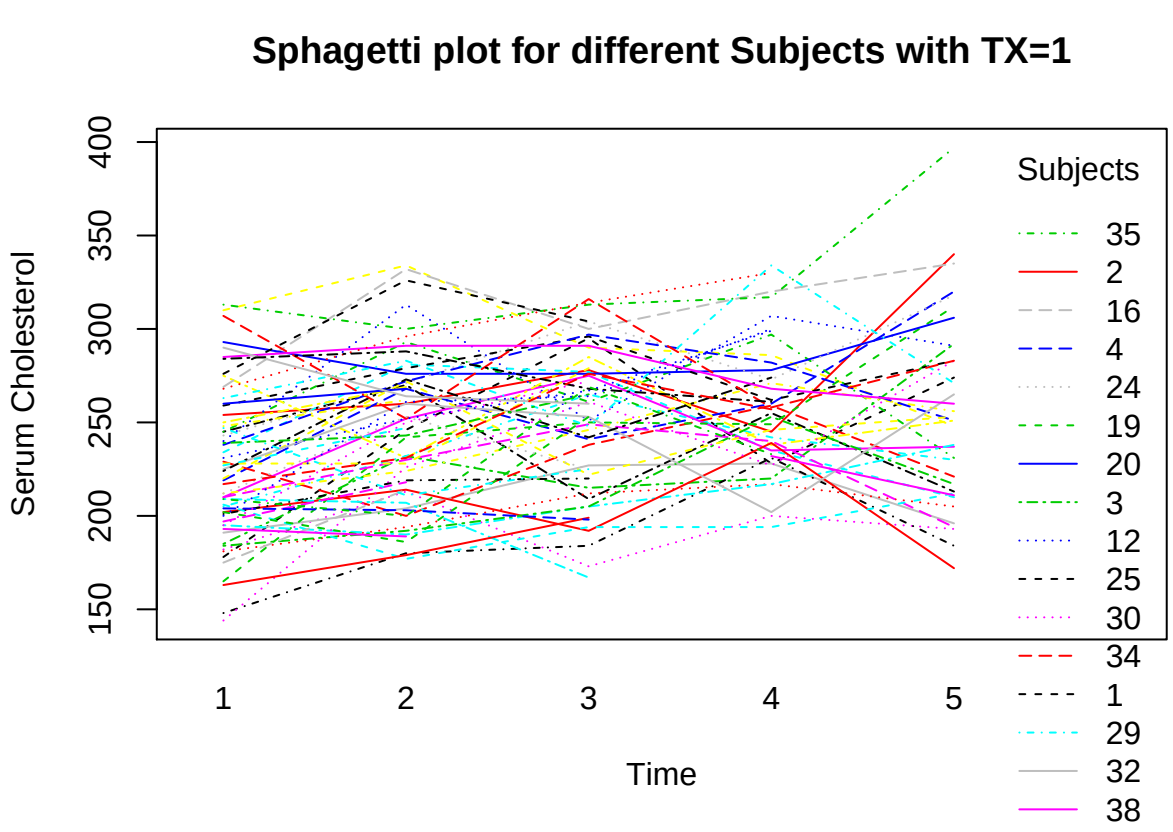
```
##
##  GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
##  gee S-function, version 4.13 modified 98/01/27 (1998)
##
## Model:
##  Link:                      Identity
##  Variance to Mean Relation: Gaussian
##  Correlation Structure:     AR-M , M = 1
##
## Call:
## gee(formula = Strength ~ Time + Treatment, id = ID, data = dta_new,
##     corstr = "AR-M")
##
## Summary of Residuals:
##        Min         1Q     Median         3Q        Max
## -7.3687246 -2.2696096  0.1021921  2.3044947  9.1603587
##
##
## Coefficients:
##               Estimate Naive S.E.  Naive z Robust S.E.  Robust z
## (Intercept) 77.8240559 1.73701919 44.80322  1.77141849 43.933185
## Time         0.2645416 0.06368252  4.15407  0.07236149  3.655834
## Treatment    1.6109803 1.04508678  1.54148  1.05636934  1.525016
##
## Estimated Scale Parameter:  11.24525
## Number of Iterations:  3
##
## Working Correlation
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 1.0000000 0.9523378 0.9069474 0.8637203 0.8225535 0.7833488 0.7460127
## [2,] 0.9523378 1.0000000 0.9523378 0.9069474 0.8637203 0.8225535 0.7833488
## [3,] 0.9069474 0.9523378 1.0000000 0.9523378 0.9069474 0.8637203 0.8225535
## [4,] 0.8637203 0.9069474 0.9523378 1.0000000 0.9523378 0.9069474 0.8637203
## [5,] 0.8225535 0.8637203 0.9069474 0.9523378 1.0000000 0.9523378 0.9069474
## [6,] 0.7833488 0.8225535 0.8637203 0.9069474 0.9523378 1.0000000 0.9523378
## [7,] 0.7460127 0.7833488 0.8225535 0.8637203 0.9069474 0.9523378 1.0000000
```

2 (a)

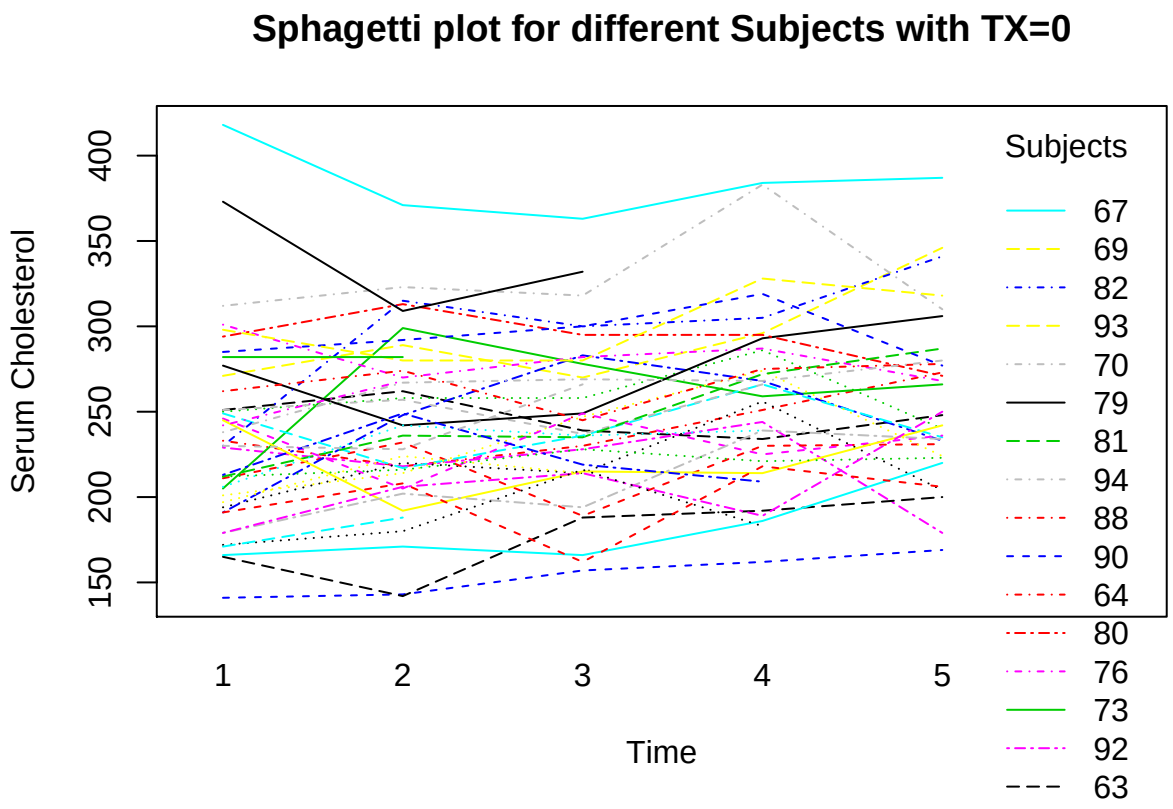
```
setwd("~/")
cholesterolData <- read.table("cholesterol.txt", na.strings = ".", header=FALSE)
names(cholesterolData) <- c("Treatment","ID","1","2","3","4","5")
cholesterolData_long <- reshape(cholesterolData, varying = list(3:7),direction = "long")
names(cholesterolData_long) <- c("Treatment", "ID", "Time", "SerumChol", "id")
attach(cholesterolData_long)
```

```
## The following objects are masked from exDataNARM_long_q2:
##
##     ID, Time, Treatment
##
## The following objects are masked from exDataNARM_long_q1:
##
##     ID, Time, Treatment
```

```
cholesterolData_long$SerumChol <- as.numeric(as.character(cholesterolData_long$SerumChol))
cholesterolData_long$Time <- factor(c("1","2","3","4","5"))
Treatment[Treatment=="2"] <- 0
interaction.plot(Time[Treatment==1], id[Treatment==1], SerumChol[Treatment==1], xlab = "Time", ylab = "
```



```
interaction.plot(Time[Treatment==0], id[Treatment==0], SerumChol[Treatment==0], xlab = "Time", ylab = "S
```



(b)

```
library(geepack)
library(splines)
dta_new <- NULL

for(i in 1:nrow(cholesterolData_long))
  dta_new <- rbind(dta_new, cholesterolData_long[cholesterolData_long$ID == i, ])

fit_geeglm <- geeglm(SerumChol ~ Treatment*bs(Time, degree=3), id=id, corstr="independence", data=choles
summary(fit_geeglm)
```

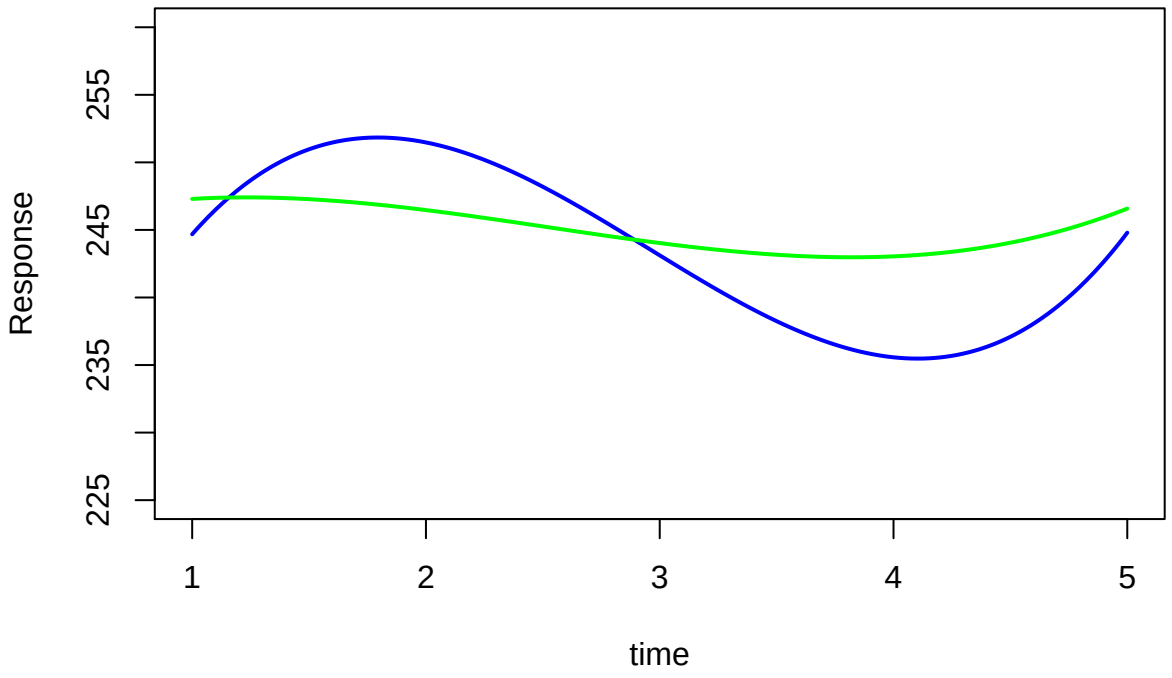
```
##
## Call:
## geeglm(formula = SerumChol ~ Treatment * bs(Time, degree = 3),
##     data = cholesterolData_long, id = id, corstr = "independence")
##
##  Coefficients:
##                                  Estimate  Std.err    Wald Pr(>|W|)
## (Intercept)                     244.6809  14.0689 302.470   <2e-16 ***
## Treatment                         2.6138  10.2191   0.065    0.798
## bs(Time, degree = 3)1            26.2518  47.4343   0.306    0.580
## bs(Time, degree = 3)2           -30.4556  40.3832   0.569    0.451
## bs(Time, degree = 3)3             0.1152  20.8816   0.000    0.996
## Treatment:bs(Time, degree = 3)1 -24.9156  34.0917   0.534    0.465
```

```
## Treatment:bs(Time, degree = 3)2  20.6497  29.3046   0.497    0.481
## Treatment:bs(Time, degree = 3)3  -0.8317  14.5627   0.003    0.954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Estimated Scale Parameters:
##             Estimate Std.err
## (Intercept)     2061   152.1
##
## Correlation: Structure = independenceNumber of clusters:   447   Maximum cluster size: 1
```

(c)

```
sq = seq(1,5,length=1000)
y_p = cbind(1,0,bs(sq,degree=3)) %*% coef(fit_geeglm)[1:5]
y_t = cbind(1,1,bs(sq,degree=3),bs(sq,degree=3)) %*% coef(fit_geeglm)[1:8]

plot(c(1, 5), c(225,260), xlab = "time", ylab = "Response", type = "n")
points(cholesterolData_long$time, cholesterolData_long$response,
       ylim=range(c(225,260)), pch = 20, col = "grey")
lines(sq, y_p, lwd = 2, col = "blue")
lines(sq, y_t, lwd = 2, col = "green")
```



(d)

```
fit_geeglm_tx <- geeglm(SerumChol ~ Treatment*bs(Time, degree=3), id=id, corstr="independence", data=cho
fit_geeglm_notx <- geeglm(SerumChol ~ bs(Time, degree=3), id=id, corstr="independence", data=cholesterol
anova(fit_geeglm_notx,fit_geeglm_tx)
```

```
## Analysis of 'Wald statistic' Table
```

```
##
## Model 1 SerumChol ~ Treatment * bs(Time, degree = 3)
## Model 2 SerumChol ~ bs(Time, degree = 3)
##   Df    X2 P(>|Chi|)
## 1  4 0.899      0.92
```

Since the p-value is 0.92 which is more than the significance level 0f 0.05. we fail to reject the null hypothesis and conclude that there is no treatment effect.