Case Studies for Note 2

Keunbaik Lee
Sungkyunkwan University

keunbaik@skku.edu

August 25, 2021

Treatment of Lead-Exposed Children (TLC) Trial I

Exposure of lead, often due to deteriorating lead-based paint in older homes, can damage cognitive function, especially in children. The CDC has decided that children with blood lead level over 10 $\mu g/dL$ are at risk.
Chelating agents can be used to treat lead poisoning, which were usually introduced by injection and required hospitalization. A new agent, succimer, can be given orally. In 1990, the *Treatment of Lead-Exposed Children (TLC) Trial Group* conducted a placebo-controlled, randomized trial of succimer in children with blood lead levels of 20-44 $\mu g/dL$. The children in the study were aged 12-33 months at enrollment. They received up to three 26-day courses of succimer or placebo and were followed for 3 years.
The data set we will look at were a random sample of 100 children, with blood levels measured at baseline, week 1, 4 and 6.

Question of Interest: whether succimer reduces blood lead levels over time relative to placebo.

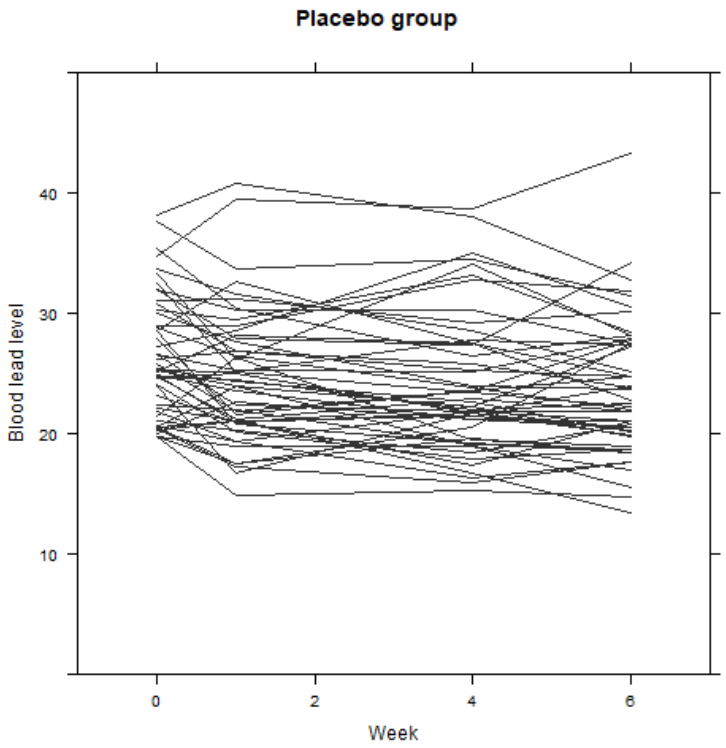
Treatment of Lead-Exposed Children (TLC) Trial II

Table: Blood lead levels ($\mu g/dL$) at baseline, week 1, 4 and 6 for 10 children in the TLC trial

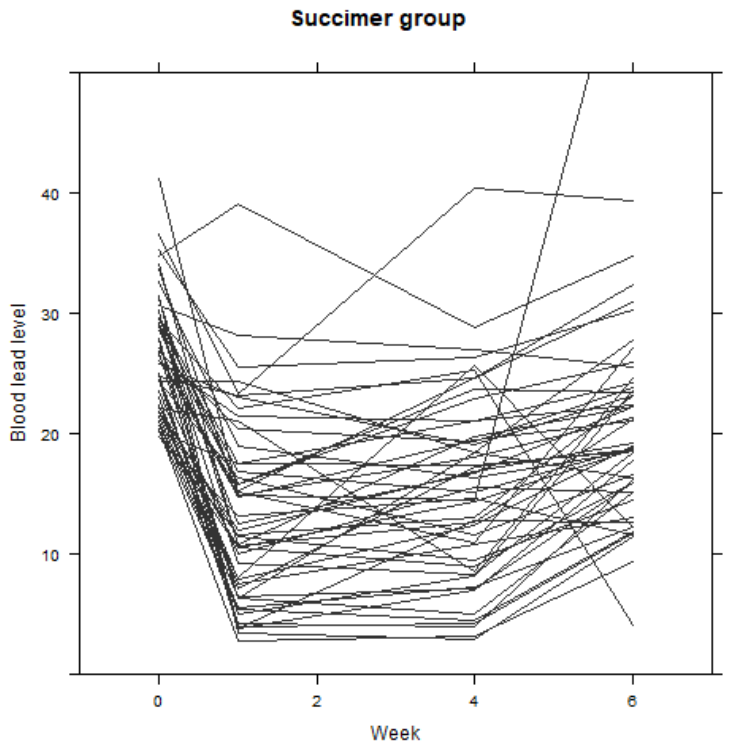
ID	Group	Baseline	Week 1	Week 2	Week 3
1	P	30.8	26.9	25.8	23.8
2	A	26.5	14.8	19.5	21.0
3	A	25.8	23.0	19.1	23.2
4	P	24.7	24.5	22.0	22.5
5	A	20.4	2.8	3.2	9.4
6	A	20.4	5.4	4.5	11.9
7	P	28.6	20.8	19.2	18.4
8	P	33.7	31.6	28.5	25.1
9	P	19.7	14.9	15.3	14.7
10	P	31.1	31.2	29.2	30.1

```
> library(lattice)
> ## Read data and compute some summary statistics
> tlc <- read.table ("tlc.txt",col.names = c("ID", "Group", "week.0",
+                          "week.1", "week.4", "week.6"))
> tlc[1:4,]
  ID Group week.0 week.1 week.4 week.6
1  1     P   30.8   26.9   25.8   23.8
2  2     A   26.5   14.8   19.5   21.0
3  3     A   25.8   23.0   19.1   23.2
4  4     P   24.7   24.5   22.0   22.5
> do.call("rbind", tapply(tlc$week.0, tlc$Group, summary))
  Min. 1st Qu. Median   Mean 3rd Qu. Max.
A 19.7  22.125  26.20 26.540  29.550 41.1
P 19.7  21.875  25.25 26.272  29.725 38.1
> by(as.matrix(tlc[,-c(1:2)]), tlc$Group,
+     function(x){
+         cbind(mean=apply(x,2,mean), sd=apply(x,2,sd))
+     }
+ )
INDICES: A
          mean       sd
week.0 26.540 5.020936
week.1 13.522 7.672487
week.4 15.514 7.852207
week.6 20.762 9.246332
------------------------------------------------------------
```

```
INDICES: P
          mean       sd
week.0 26.272 5.024107
week.1 24.660 5.461180
week.4 24.070 5.753127
week.6 23.646 5.639808
```



(a) Placebo



(b) Succimer

Figure: Plot of blood lead levels by treatment group

Notes

- Complete and balanced data.
- Interested in marginal inference: i.e., compare the mean profiles of the two groups over time.
- Randomized trial.
- The mean profile does not appeat to be linear, especially for the treatment group.

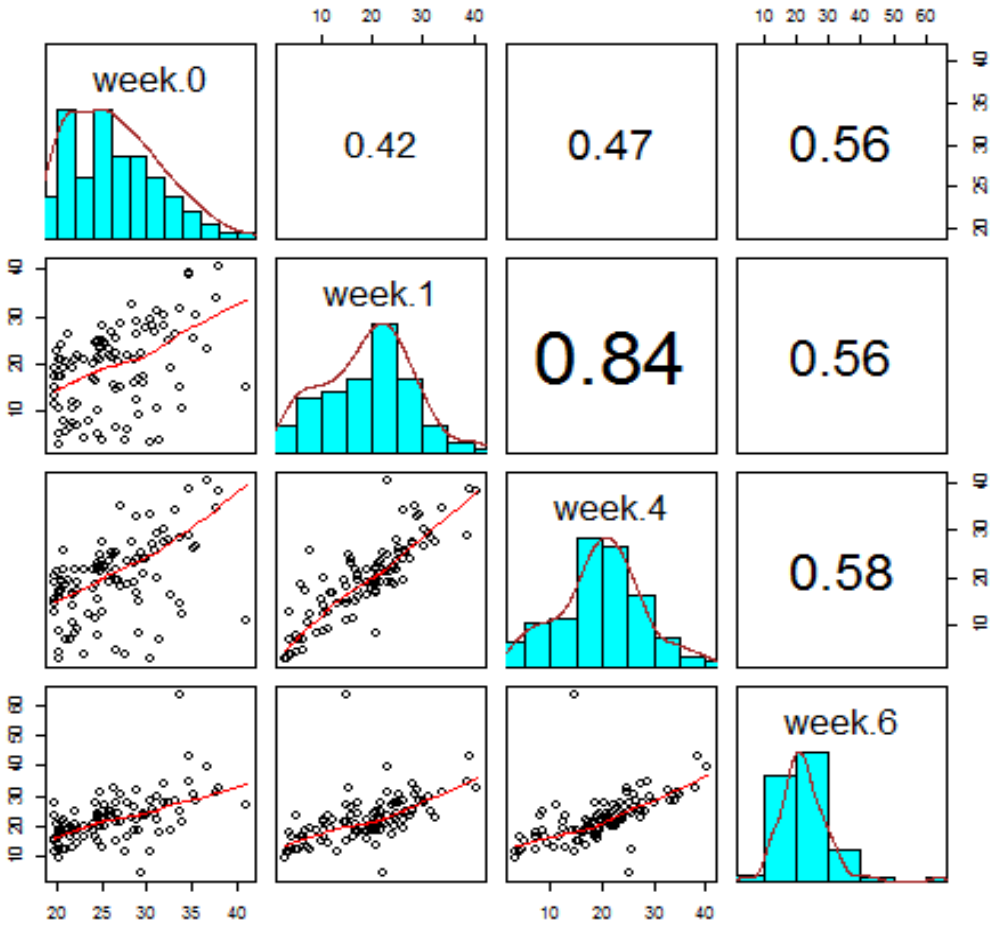


Figure: Pairwise scatter-plot of blood lead levels at baseline, week 1, 4, and 6 for children in TLC trial.

R cdoes

```
library(lattice)
## Read data and compute some summary statistics
tlc <- read.table ("tlc.txt",col.names = c("ID", "Group", "week.0",
                        "week.1", "week.4", "week.6"))
tlc[1:4,]
do.call("rbind", tapply(tlc$week.0, tlc$Group, summary))
by(as.matrix(tlc[,-c(1:2)]), tlc$Group,
    function(x){cbind(mean=apply(x,2,mean), sd=apply(x,2,sd))}
)

## Explore data
tlcL <- reshape (tlc, direction = "long", idvar = "ID", varying = 3:6)
tlcL[95:105,]
names(tlcL)[3:4] <- c("Week", "Lead")
# Scatterplot by treatment group with LOESS smoothing curve
xyplot (Lead ~ Week | Group, data = tlcL,
groups = tlcL$ID, type = "l",
panel = function(x, y, subscripts, groups, ...) {
panel.superpose (x, y,
panel.groups = "panel.xyplot",
subscripts,
groups, col = "gray40", ...)
panel.loess(x, y, col = "red", lwd = 2, ...)
})
```

Objectives of Analysis I

The null hypothesis of no treatment effect can be expressed in different ways:

- $H_0 : \mu_j(A) = \mu_j(P)$ for all $j = 1, 2, 3, 4$.
  - Time is treated as a factor.
  - This null can be expressed in terms of both the regression coefficients for the treatment and time $\times$ treatment interactions.
- $H_0 : \mu_j(A) - \mu_1(A) = \mu_j(P) - \mu_1(P)$ for all $j = 1, 2, 3, 4$.
  - Emphasis on the treatment effect on the **changes**, i.e., time$\times$treatment interaction.
  - Less restrictive, allows the baseline lead levels to differ between groups.
- Model the response profile via a parametric (or non-parametric) model, i.e., a linear or quadratic model, and test the time$\times$treatment interaction.
  - Linear model is not appropriate.

```
> # Simple linear regression
> # make referecne category for "P"
> tlcl$Group.f <- relevel(factor(tlcl$Group),ref="P")
> tlc.lm <- lm(lead ~ factor(week)*Group.f, data=tlcl)
> summary(tlc.lm)

Call:
lm(formula = lead ~ factor(week) * Group.f, data = tlcl)

Residuals:
    Min      1Q  Median      3Q     Max
-16.662  -4.620  -0.993   3.673  43.138

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 26.272      0.937  28.038  < 2e-16 ***
factor(week)1               -1.612      1.325  -1.216   0.2245
factor(week)4               -2.202      1.325  -1.662   0.0974 .
factor(week)6               -2.626      1.325  -1.982   0.0482 *
Group.fA                     0.268      1.325   0.202   0.8398
factor(week)1:Group.fA     -11.406      1.874  -6.086 2.75e-09 ***
factor(week)4:Group.fA      -8.824      1.874  -4.709 3.47e-06 ***
factor(week)6:Group.fA      -3.152      1.874  -1.682   0.0934 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6.626 on 392 degrees of freedom
Multiple R-squared:  0.3284,    Adjusted R-squared:  0.3164
F-statistic: 27.38 on 7 and 392 DF,  p-value: < 2.2e-16

> anova(tlc.lm)
Analysis of Variance Table

Response: lead
                      Df  Sum Sq Mean Sq F value    Pr(>F)
factor(week)           3  3272.8  1090.9  24.850 9.701e-15 ***
Group.f                1  3110.9  3110.8  70.862 7.281e-16 ***
factor(week):Group.f   3  2030.4   676.8  15.417 1.685e-09 ***
Residuals            392 17208.8    43.9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R cdoes

```
## Simple Linear Model
temp <- lm(Lead ~ factor (Week) * Group, data = tlcL)
summary (temp)
anova (temp)
# Model diagnosis
par(mfrow=c(2,2))
plot(temp)
```

GEE I

- In R, GEE(for linear model, it just means robust variance estimation) is implemented by libraries **gee** and a newer geepack (the function name is geese).
- Note that it is necessary to sort the data by ID first.
- By default, gee uses “working independence” correlation matrix.
- Results:
  - The “naive” SEs are based on the specified correlation matrix (what we called “model-based” SEs). Note that here they are the same as in the simple linear model.
  - The coefficients are the same as in OLS.
  - The robust estimates of SE are smaller (more efficient).
  - There appears to be an outlier but we will ignore it.
  - Since GEE is not based on likelihood, we cannot use likelihood ratio or score tests. We can use Wald test to test the null hupothesis of no Week:Group interaction effect but some programs seems necessary.

GEE II

- temp$robust.variance gives the full covariance matrix for β.
- The robust standard error estimates are same for different correlation models.

```
> # General linear model using GEE
> library(gee)
> # independence working correlation
> tlcl.gee.indep <- gee(lead ~ factor(week)*Group.f,,id=ID,corstr="independence",data=tlcl)
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
           (Intercept)          factor(week)1          factor(week)4
                26.272                 -1.612                 -2.202
         factor(week)6               Group.fA factor(week)1:Group.fA
                -2.626                  0.268                -11.406
factor(week)4:Group.fA factor(week)6:Group.fA
                -8.824                 -3.152
> summary(tlcl.gee.indep)

 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:     Independent

Call:
gee(formula = lead ~ factor(week) * Group.f, id = ID, data = tlcl,
    corstr = "independence")

Summary of Residuals:
     Min       1Q   Median       3Q      Max
-16.6620  -4.6205  -0.9930   3.6725  43.1380


Coefficients:
                     Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept)            26.272  0.9370175 28.0378980   0.7033749  37.3513444
factor(week)1          -1.612  1.3251428 -1.2164727   0.4330325  -3.7225846
```

```
factor(week)4            -2.202  1.3251428 -1.6617077   0.4386752  -5.0196593
factor(week)6            -2.626  1.3251428 -1.9816732   0.5278091  -4.9752834
Group.fA                  0.268  1.3251428  0.2022424   0.9944085   0.2695069
factor(week)1:Group.fA  -11.406  1.8740349 -6.0863327   1.1086833 -10.2878794
factor(week)4:Group.fA   -8.824  1.8740349 -4.7085569   1.1408849  -7.7343471
factor(week)6:Group.fA   -3.152  1.8740349 -1.6819324   1.2439296  -2.5339055

Estimated Scale Parameter:  43.90009
Number of Iterations:  1

Working Correlation
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
>
> # exchangeable working correlation
> tlcl.gee.exchange <- gee(lead ~ factor(week)*Group.f,,id=ID,corstr="exchangeable",data=tlcl
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
           (Intercept)          factor(week)1          factor(week)4
                26.272                 -1.612                 -2.202
         factor(week)6               Group.fA factor(week)1:Group.fA
                -2.626                  0.268                -11.406
factor(week)4:Group.fA factor(week)6:Group.fA
                -8.824                 -3.152
> summary(tlcl.gee.exchange)

 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:     Exchangeable

Call:
gee(formula = lead ~ factor(week) * Group.f, id = ID, data = tlcl,
    corstr = "exchangeable")

Summary of Residuals:
     Min       1Q   Median       3Q      Max
-16.6620  -4.6205  -0.9930   3.6725  43.1380


Coefficients:
                       Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept)              26.272  0.9370175 28.0378980   0.7033749  37.3513444
factor(week)1            -1.612  0.8470380 -1.9031023   0.4330325  -3.7225846
factor(week)4            -2.202  0.8470380 -2.5996472   0.4386752  -5.0196593
factor(week)6            -2.626  0.8470380 -3.1002150   0.5278091  -4.9752834
Group.fA                  0.268  1.3251428  0.2022424   0.9944085   0.2695069
factor(week)1:Group.fA  -11.406  1.1978927 -9.5217212   1.1086833 -10.2878794
factor(week)4:Group.fA   -8.824  1.1978927 -7.3662693   1.1408849  -7.7343471
factor(week)6:Group.fA   -3.152  1.1978927 -2.6312875   1.2439296  -2.5339055

Estimated Scale Parameter:  43.90009
Number of Iterations:  1

Working Correlation
          [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.5914168 0.5914168 0.5914168
[2,] 0.5914168 1.0000000 0.5914168 0.5914168
[3,] 0.5914168 0.5914168 1.0000000 0.5914168
[4,] 0.5914168 0.5914168 0.5914168 1.0000000
>
> # unstructured working correlation
> tlcl.gee.unstruct <- gee(lead ~ factor(week)*Group.f,,id=ID,corstr="unstructured",data=tlcl
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
running glm to get initial regression estimate
           (Intercept)          factor(week)1          factor(week)4
                26.272                 -1.612                 -2.202
         factor(week)6               Group.fA factor(week)1:Group.fA
                -2.626                  0.268                -11.406
factor(week)4:Group.fA factor(week)6:Group.fA
                -8.824                 -3.152
> summary(tlcl.gee.unstruct)

 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:     Unstructured

Call:
gee(formula = lead ~ factor(week) * Group.f, id = ID, data = tlcl,
    corstr = "unstructured")

Summary of Residuals:
     Min       1Q   Median       3Q      Max
-16.6620  -4.6205  -0.9930   3.6725  43.1380


Coefficients:
                      Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept)             26.272  0.9370175 28.0378980   0.7033749  37.3513444
factor(week)1           -1.612  0.9958441 -1.6187273   0.4330325  -3.7225846
factor(week)4           -2.202  0.9838820 -2.2380732   0.4386752  -5.0196593
factor(week)6           -2.626  0.9316319 -2.8187099   0.5278091  -4.9752834
Group.fA                 0.268  1.3251428  0.2022424   0.9944085   0.2695069
factor(week)1:Group.fA -11.406  1.4083362 -8.0989182   1.1086833 -10.2878794
factor(week)4:Group.fA  -8.824  1.3914193 -6.3417260   1.1408849  -7.7343471
```

```
factor(week)6:Group.fA   -3.152  1.3175265 -2.3923618   1.2439296  -2.5339055

Estimated Scale Parameter:  43.90009
Number of Iterations:  1

Working Correlation
          [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.4352486 0.4487346 0.5057311
[2,] 0.4352486 1.0000000 0.8094551 0.6759677
[3,] 0.4487346 0.8094551 1.0000000 0.6975035
[4,] 0.5057311 0.6759677 0.6975035 1.0000000
```

```
## GEE
library (gee)
tlcL <- tlcL[order(tlcL$Group,tlcL$ID,tlcL$Week),] # sorting by group,

# default(independence working correlation)
temp <- gee (Lead ~ factor (Week) * Group, id = ID, data = tlcL)
summary (temp)

# exchangeable working correlation
temp <- gee (Lead ~ factor (Week) * Group, id = ID,
 corstr = "exchangeable", data = tlcL)
summary (temp)

# unstructured working correlation
temp <- gee (Lead ~ factor (Week) * Group, id = ID,
 corstr = "unstructured", data = tlcL)
summary (temp)
```

Generalized Least Squares I

- R library nlme provides a function gls that does generalized least squares estimation.
- The difference with gee is that is does not compute sandwich standard error estimates.
- By default, REML is used. We requested maximum likelihood by specifying the method argument. In this case, there is very little difference.
- Since REML is “conditional” on the fixed effects, when comparing models with different fixed effects (regression coefficients), maximum likelihood should be used.
- gls does anova (F-test).

```
> # General linear model using ML & REML
> library(nlme)
>
> # sorting data by ID and week
> o <- order(tlcl$Group,tlcl$ID,tlcl$week)
> tlcl <- tlcl[o,]
> # ML
> tlc.ml <- gls(lead~factor(week)*Group.f, data=tlcl, method="ML",
+                   correlation=corCompSymm(form=~1|ID))
> summary(tlc.ml)
Generalized least squares fit by maximum likelihood
  Model: lead ~ factor(week) * Group.f
  Data: tlcl
       AIC      BIC    logLik
  2490.822 2530.736 -1235.411

Correlation Structure: Compound symmetry
 Formula: ~1 | ID
 Parameter estimate(s):
      Rho
0.5954401

Coefficients:
                          Value Std.Error   t-value p-value
(Intercept)              26.272 0.9370175 28.037898  0.0000
factor(week)1            -1.612 0.8428574 -1.912542  0.0565
factor(week)4            -2.202 0.8428574 -2.612542  0.0093
factor(week)6            -2.626 0.8428574 -3.115592  0.0020
Group.fA                  0.268 1.3251428  0.202242  0.8398
factor(week)1:Group.fA -11.406 1.1919804 -9.568950  0.0000
factor(week)4:Group.fA  -8.824 1.1919804 -7.402807  0.0000
factor(week)6:Group.fA  -3.152 1.1919804 -2.644339  0.0085

 Correlation:
                       (Intr) fct()1 fct()4 fct()6 Grp.fA f()1:G f()4:G
```

```
factor(week)1          -0.450
factor(week)4          -0.450  0.500
factor(week)6          -0.450  0.500  0.500
Group.fA               -0.707  0.318  0.318  0.318
factor(week)1:Group.fA  0.318 -0.707 -0.354 -0.354 -0.450
factor(week)4:Group.fA  0.318 -0.354 -0.707 -0.354 -0.450  0.500
factor(week)6:Group.fA  0.318 -0.354 -0.354 -0.707 -0.450  0.500  0.500

Standardized residuals:
       Min         Q1        Med         Q3        Max
-2.5402789 -0.7044388 -0.1513922  0.5599072  6.5767945

Residual standard error: 6.559122
Degrees of freedom: 400 total; 392 residual
> # REML
> tlc.reml <- gls(lead~factor(week)*Group.f, data=tlcl, method="REML",
+                 correlation=corCompSymm(form=~1|ID))
> summary(tlc.reml)
Generalized least squares fit by REML
  Model: lead ~ factor(week) * Group.f
  Data: tlcl
       AIC      BIC    logLik
  2480.621 2520.334 -1230.311

Correlation Structure: Compound symmetry
 Formula: ~1 | ID
 Parameter estimate(s):
      Rho
0.5954401

Coefficients:
                        Value Std.Error   t-value p-value
(Intercept)            26.272 0.9370175 28.037898  0.0000
factor(week)1          -1.612 0.8428574 -1.912542  0.0565
factor(week)4          -2.202 0.8428574 -2.612542  0.0093
```

```
factor(week)6           -2.626 0.8428574 -3.115592  0.0020
Group.fA                 0.268 1.3251428  0.202242  0.8398
factor(week)1:Group.fA -11.406 1.1919804 -9.568950  0.0000
factor(week)4:Group.fA  -8.824 1.1919804 -7.402807  0.0000
factor(week)6:Group.fA  -3.152 1.1919804 -2.644339  0.0085

 Correlation:
                      (Intr) fct()1 fct()4 fct()6 Grp.fA f()1:G f()4:G
factor(week)1         -0.450
factor(week)4         -0.450  0.500
factor(week)6         -0.450  0.500  0.500
Group.fA              -0.707  0.318  0.318  0.318
factor(week)1:Group.fA  0.318 -0.707 -0.354 -0.354 -0.450
factor(week)4:Group.fA  0.318 -0.354 -0.707 -0.354 -0.450  0.500
factor(week)6:Group.fA  0.318 -0.354 -0.354 -0.707 -0.450  0.500  0.500

Standardized residuals:
       Min         Q1        Med         Q3        Max
-2.5147478 -0.6973588 -0.1498706  0.5542799  6.5106944

Residual standard error: 6.625714
Degrees of freedom: 400 total; 392 residual
```

```
## Generalized Least Squares
library(nlme)
# ML method
temp <- gls (Lead ~ factor (Week) * Group, data = tlcL, method = "ML",
 correlation = corCompSymm (form = ~ 1 | ID))
anova (temp)
intervals (temp)

# REML method
temp <- gls (Lead ~ factor (Week) * Group, data = tlcL,
 correlation = corCompSymm (form = ~ 1 | ID))
anova (temp)
intervals (temp)
```

Dealing with Baseline Outcome I

When only two measurements are taken for each subject, say pre- and post-treatments (Y_{i0} and Y_{i1}). Consider the three possible models:

$$
\begin{aligned}
Y_{i1} &= \mu + \beta_1 X_i + \epsilon_i && (1)\\
(Y_{i1} - Y_{i0}) &= \mu^* + \beta_1^* X_i + \epsilon_i && (2)\\
Y_{i1} &= \mu^{**} + \beta_1^{**} X_i + \beta_2 Y_{i0} + \epsilon_i && (3)
\end{aligned}
$$

- For randomized trials, it can be shown that $\beta_1 = \beta_1^* = \beta_1^{**}$. The last two models may be more precise.
- For observational studies, the "post-only" model () is generally not satisfactory. The "change" model (**??**) and the "adjust" model (**??**) have different interpretations and often quite different values for β_1.

```
# Dealing with baseline outcome
out1<-lm(week.1~Group, data=tlc)
summary(out1)

out1.0<-lm(I(week.1 - week.0) ~ Group, data = tlc)
summary(out1.0)

out1.grp.week0 <-lm(week.1 ~ Group + week.0, data = tlc)
summary(out1.grp.week0)
```

Baseline Response for Longitudinal Data I

In the case where more than two observations ("waves") are taken, consider the four ways of handling the baseline value:

1. Retain it as part of the outcome vector and make no assumptions about group differences in the mean response at baseline.
2. Retain it as part of the outcome and assume the group means are equal at baseline, such as in a randomized trial.
3. Subtract the baseline response from all remaining responses.
4. Use the baseline value as a covariate in the analysis.

- In methods 1 and 2, the null hypothesis is that the Group by Week interaction effects are zero.
- There is no Group main effect in method 2. When it is appropriate, method2 is more powerful than method 1.

Baseline Response for Longitudinal Data II

- In methods 3 and 4, the null hypothesis is that both the Group main effect and Group by Week interaction effects are zero.
- Method 4 is more powerful than method 3.
- Methods 1 and 3 are equivalent with method 1 being more powerful.
- Methods 2 and 4 are similar with method 2 being more powerful.
- Methods 2 and 4 are only appropriate when it is reasonable to assume the baseline means are equal between groups (for randomized trial) or can be (conceptually at least) "held" equal between groups (for observational studies).

```
# Method 1
full.1 <- gls (Lead ~ factor (Week) * Group, method = "ML", data = tlcL,
  correlation = corCompSymm (form = ~ 1 | ID))
reduced.1 <- gls (Lead ~ factor (Week) + Group, method = "ML", data = tlcL,
  correlation = corCompSymm (form = ~ 1 | ID))
anova (full.1, reduced.1)

# Method 2
tlcL$W1P <- (tlcL$Week == 1) & (tlcL$Group == "P")
tlcL$W4P <- (tlcL$Week == 4) & (tlcL$Group == "P")
tlcL$W6P <- (tlcL$Week == 6) & (tlcL$Group == "P")
full.2 <- gls (Lead ~ factor (Week) + W1P + W4P + W6P, data = tlcL, method = "ML",
  correlation = corCompSymm (form = ~ 1 | ID))
reduced.2 <- gls (Lead ~ factor (Week), data = tlcL, method = "ML",
  correlation = corCompSymm (form = ~ 1 | ID))
anova (full.2, reduced.2)

# Method 3
tlcL2 <- reshape (tlc, direction = "long", idvar = "ID", varying = 4:6)
names (tlcL2)[3:5] <- c("BaseLead", "Week", "Lead")
tlcL2$ChangeLead <- tlcL2$Lead - tlcL2$BaseLead
tlcL2 <- tlcL2[order (tlcL2$Group, tlcL2$ID, tlcL2$Week),]

full.3 <- gls (ChangeLead ~ factor (Week) * Group, method = "ML", data = tlcL2,
 correlation = corCompSymm (form = ~ 1 | ID))
reduced.3 <- gls (ChangeLead ~ factor (Week), method = "ML", data = tlcL2,
 correlation = corCompSymm (form = ~ 1 | ID))
anova (full.3, reduced.3)

# Method 4
full.4 <- gls (Lead ~ factor (Week) * Group + BaseLead, method = "ML",data = tlcL2,
  correlation = corCompSymm (form = ~ 1 | ID))
reduced.4 <- gls (Lead ~ factor (Week) + BaseLead, method = "ML", data = tlcL2,
  correlation = corCompSymm (form = ~ 1 | ID))
anova (full.4, reduced.4)
```

Inference for Marginal Mean Effects I

- Approximate Wald tests (and associated confidence intervals) can be used (with robust variance estimates if so desired).
- For small sample sizes, approximate $t-$ or $F-$tests may be more accurate. However, the estimation of the proper number of degrees of freedom is non-trivial (SAS includes four different methods in PROC MIXED).
- For nested models, likelihood ratio test can be used. However, it is not valid if the models are fitted using REML rather than ML.
- Other model selection criteria, such AIC or BIC, can be used.

Model Diagnosis I

- The model diagnosis for general linear model is similar to linear models.
- Library nlme provides several functions for examining gls objects.
- Results:
  - The errors should center at about zero and the variances should be approximately equal.
  - Variance and mean relationship: slight increase in variance with time.
  - An outliner with ID 40.
  - There are several types of residuals, **raw**, **Pearson** and **normalized**.

```
## Model Diagnosis
# Residual Plots
plot (full.1, ID ~ resid (.), id = 0.01)
plot (full.1, resid (.) ~ Week | Group, abline = 0,
 id = ~ ID == 40)
plot (full.1, resid (., type = "p") ~ fitted (.) | Group,id = ~ ID == 40) # residual plot
plot (full.1, Lead ~ fitted (.))
qqnorm (full.1, ~ resid (.)) # checking normality assumption
```