STAT320/420

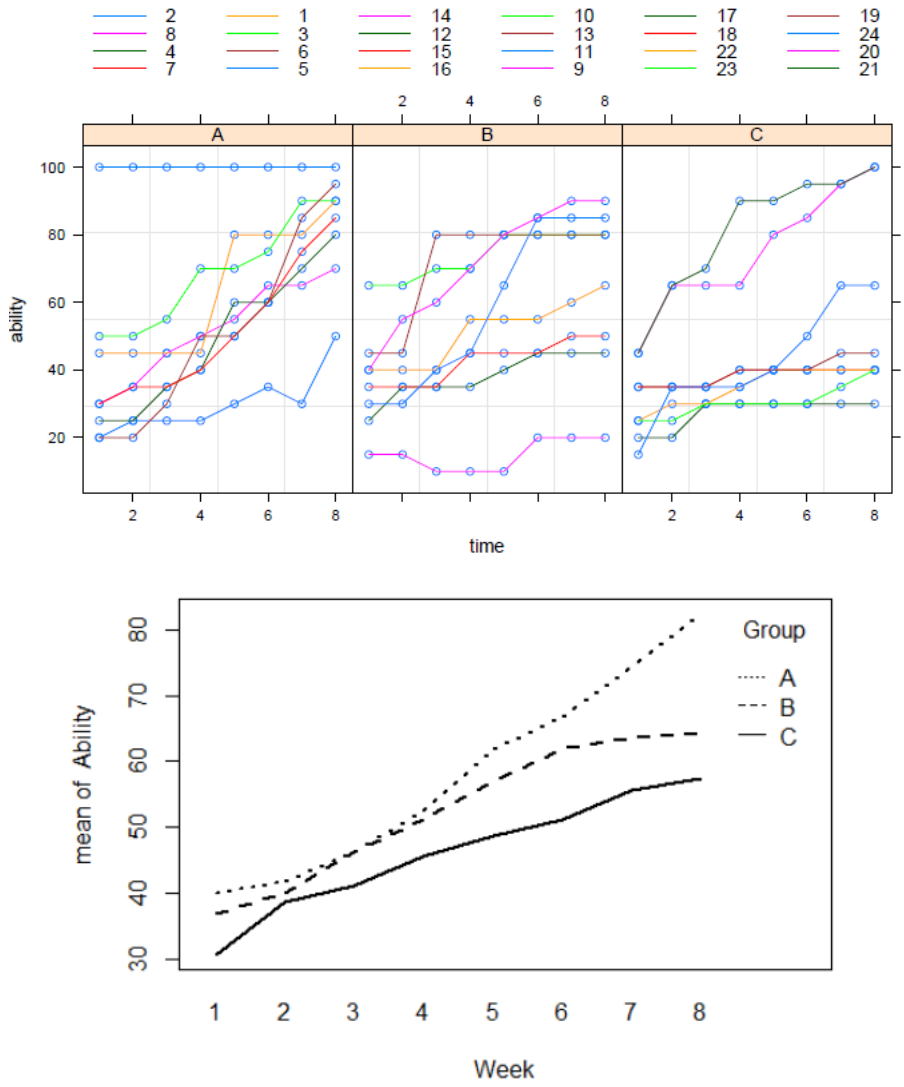

Figure 1: Top panel: Plot of ability against time for each subject by group, bottom panel: Interaction plot of Ability and Week

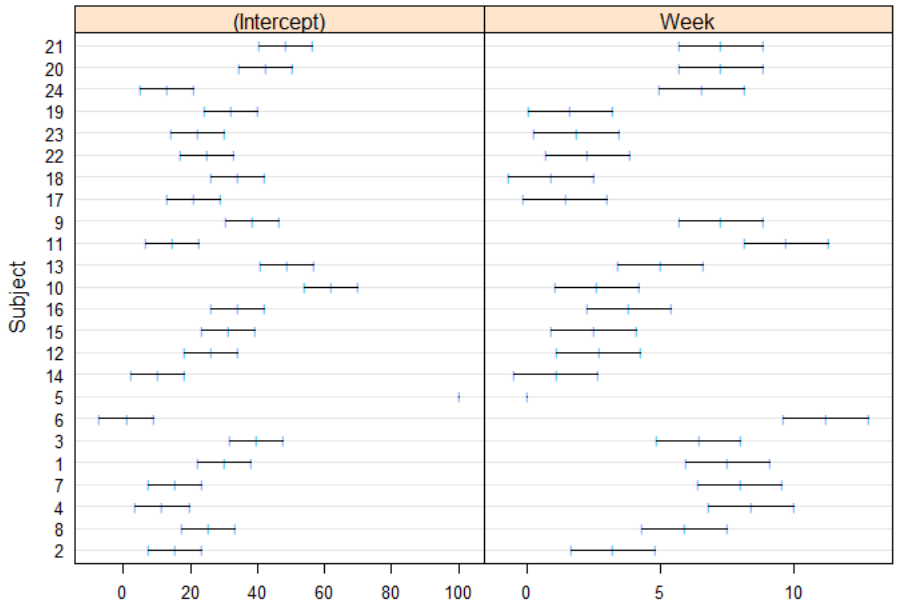

Figure 2: Confidence intervals for random intercepts and slopes by Subject

(a - b) Figure 1 - top panel shows the trace lines for each subject within each group. It is clear that the change in ability over time is not the same for all subjects within a group (i.e., slopes differ) and also that the ability at the beginning of the therapy differs among subjects (different intercepts).

Figure 1 - bottom panel shows that mean ability increases overtime (averaged over subjects) and that the change in time is similar for each group. However, at each time the mean ability of patients in the new OT therapy group (A) is the same or higher than that of the other two groups, and the difference in mean ability between A and the other two increases over time. Note that, from time 6 onwards the increase in ability is not as great for groups B and C as it is for A, which suggests that the new OT therapy (A) may be better at sustaining improvement. It seems reasonable to fit a linear trend over time while allowing for an interaction between group and time.

Figure 2 gives the CIs for the random slopes and intercepts. This also highlights the difference among intercepts and slopes for Subjects. This suggests fitting a random slopes, random intercepts repeated measures mixed effects model.

The model can be written as:

$$\begin{aligned} y_{ijk} &= \mu + \delta_i + \tau_j + (\delta:\tau)_{ij} + S_k + (S:\tau)_{ik} + \epsilon_{ijk} \\ \text{where} \quad & S \sim N(0, \sigma^2_U)\ , (S:\tau) \sim N(0, \sigma^2_{S:\tau})\ , \epsilon \sim N(0, \sigma^2)\ , \text{cov}(\epsilon, \mathbf{U}) = 0 \end{aligned}$$

δ_i represents fixed group effect; τ_j represents time (fixed); $(\delta:\tau)_{ij}$ represents group:time interaction (fixed); S represents subject (random intercepts); $(S:\tau)_{ik}$ represents random slopes.

(c) **Random intercepts only**

This model suggests a difference in the change overtime among the groups. The Group:time interaction term in Table 1 is significant, p=0.0002. The greatest increase in mean ability over time occurring with Therapy A. The coefficients for GroupB:time and GroupC:time (difference in slopes) are both negative and statistically significant. However, if you produce the ACF, you will see that there is a distinctive pattern, suggesting that correlation amongst observations has not been accounted for.

Table 1: ANOVA for fixed effects in the mixed effects model for the stroke data, allowing for random intercepts

```
mod.lme <- lme(Ability ~ Group*Week, random=~1|Subject, data=stroke2)

            numDF denDF F-value p-value
(Intercept)     1   165   158.9  <.0001
Group           2    21     0.7  0.5035
time            1   165   311.9  <.0001
Group:time      2   165     8.9  0.0002
```

Table 2: Summary of coefficients and variance components for a mixed effects analysis of the stroke data, allowing for random intercepts

```
mod.lme <- lme(Ability ~ Group*Week, random=~1|Subject, data=stroke2)

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:        20.1     8.56

Fixed effects: ability ~ Group * time
            Value Std.Error  DF t-value p-value
(Intercept) 29.82      7.50 165    3.98  0.0001
GroupB       3.35     10.60  21    0.32  0.7553
GroupC      -0.02     10.60  21    0.00  0.9983
time         6.32      0.47 165   13.54  0.0000
GroupB:time -1.99      0.66 165   -3.02  0.0030
GroupC:time -2.69      0.66 165   -4.07  0.0001
```

Mixed effects model with random slopes & intercepts

Table 3: ANOVA for fixed effects in the mixed effects model for the stroke data, allowing for random slopes & intercepts

```
mod2.lme <- lme(Ability ~ Group*Week, random = ~ Week|Subject, data=stroke2)

            numDF denDF F-value p-value
(Intercept)     1   165   121.1  <.0001
Group           2    21     0.3   0.739
time            1   165    58.3  <.0001
Group:time      2   165     1.7   0.192
```

Table 4: Summary of coefficients and variance components for a mixed effects analysis of the stroke data, allowing for random slopes & intercepts

```
mod2.lme <- lme(Ability ~ Group*Week, random = ~ Week|Subject, data=stroke2)

Random effects:
 Formula: ~time | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 21.03  (Intr)
time         2.95  -0.372
Residual     5.18

Fixed effects: ability ~ Group + time + Group:time
            Value Std.Error  DF t-value p-value
(Intercept) 29.82      7.57 165    3.94  0.0001
GroupB       3.35     10.71  21    0.31  0.7576
GroupC      -0.02     10.71  21    0.00  0.9984
time         6.32      1.08 165    5.85  0.0000
GroupB:time -1.99      1.53 165   -1.31  0.1937
GroupC:time -2.69      1.53 165   -1.76  0.0806
```

Model comparison:

From Table 5, we see that the inclusion of the random slopes term is justified in this case, with the random slopes model (mod2.lme) having a much lower AIC (1346) than mod1.lme (1453).

Table 5: Comparison of two mixed effects models for the stroke data

```
mod.lme <- lme(Ability ~ Group*Week, random=~1|Subject, data=stroke2)
mod2.lme <- lme(Ability ~ Group*Week, random = ~ Week|Subject, data=stroke2)
###
          Model df  AIC  BIC logLik   Test L.Ratio p-value
mod.lme       1  8 1453 1479   -718
mod2.lme      2 10 1346 1378   -663 1 vs 2     111  <.0001
```

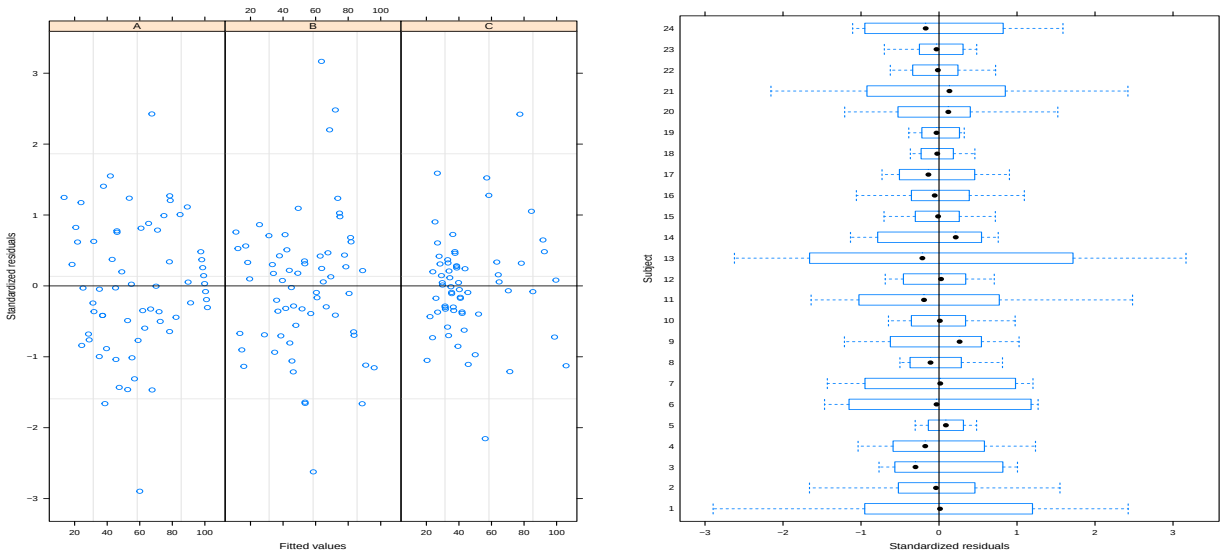

Figure 3: Diagnostic plots

(d-e-f)

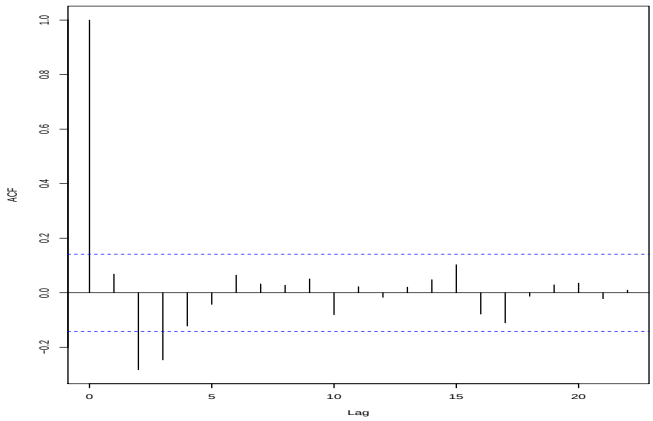

Figure 4: ACF of residuals for stroke analysis

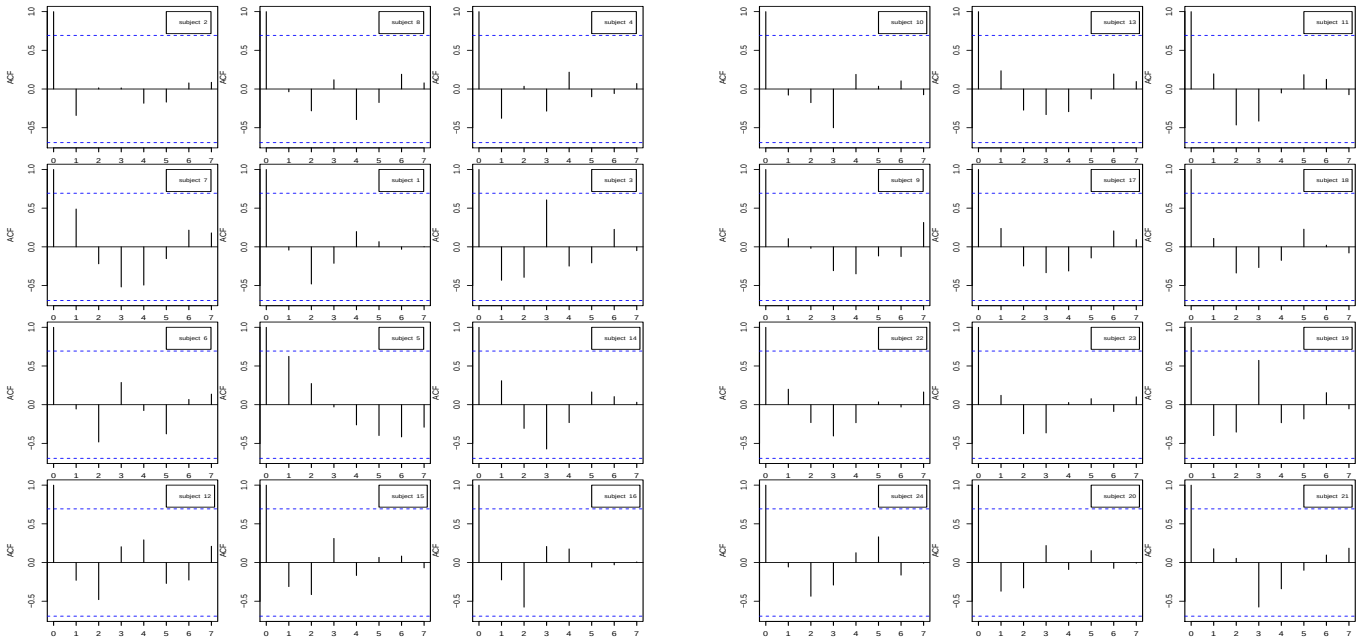

Figure 5: ACF of residuals for stroke analysis by subject

The plot on the left hand side of Figure 3 shows a random scatter of residuals around 0. However there are a few outliers with |std residuals| > 2, associated with Subjects 1, 11, 13 & 21 (refer Figure 3). There were 2 outliers in Group A, 4 in Group B and 2 in Group C. The most extreme observation is that for subject 13 (Group B) at time 3 (residual = 3.166). This subject had a much greater jump than predicted on the Bartel index from time 2 to time 3 (refer to Figure 1).

You can check the residuals by printing them out (`resid(mod2.lme,type="p")`) and matching them against the original data or you could identify them from Figure 3.

The plot on the right hand side of Figure 3 shows the distribution of residuals by subject. They are distributed around 0. There does not appear to be a problem with model assumptions, apart from the outliers previously identified.

A cyclical pattern is apparent in Figure 4, particularly for the first few lags, with lags 3 and 4 extending beyond the bands, suggesting that the residuals are correlated and a correlation structure needs to be implemented. This pattern is also apparent in the acf for some subjects (e.g., Subjects 7 and 17).

(g) Figure 6 shows that the increase in mean ability overtime for Groups B and C are similar (parallel lines). The increase in mean ability overtime seems greater for treatment A, but we will see in (vii) below that the difference is not significant.

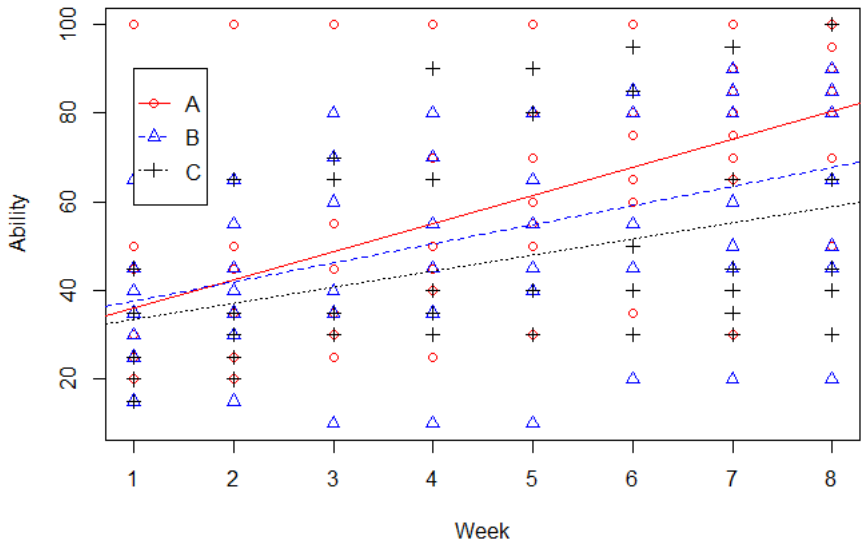

Figure 6: Fitted lines

(h) From Table 4, we can obtain the regression coefficients for the fixed effects. Group A is the reference group.

Group A

Mean ability = 29.82 + 6.32 × time

Group B

Mean ability = (29.82 +3.35) + (6.32 - 1.99) × time = 33.17 + 4.33 × time.

Group C

Mean ability = (29.82 -0.02) + (6.32 - 2.69) × time = 29.8 + 3.63 × time.

From Tables 3 & 4 we can see that there does not appear to be significant difference among treatments. The interaction term in Table 1 is not significant (p= 0.19). The slopes for Groups B and C are not significantly different from that of A (Table 4: p-values = 0.219

and 0.08, respectively). There is no need to reset the reference group here as A is the value with the greatest slope, so if B and C are not significantly different from A, they will also not be significantly different from each other.

If you want, you can refit the model without the interaction (mod3.lme) and model with only Week in the fixed effect (mod4.lme), then compare the 3 models.

Table 6: Comparison of fixed effects in a mixed model using the ML method

```
mod2.lme <- lme(Ability~Group*Week,random =~Week|Subject,data=stroke2,method="ML")
mod3.lme <- lme(Ability~Group+Week,random =~Week|Subject,data=stroke2,method="ML")
mod4.lme <- lme(Ability~Week,random =~Week|Subject,data=stroke2,method="ML")
anova(mod2.lme,mod3.lme,mod4.lme)

       Model df  AIC  BIC logLik   Test L.Ratio p-value
fit1ml     1 10 1368 1401 -674.2
fit2ml     2  8 1368 1394 -676.0 1 vs 2   3.535  0.1708
fit3ml     3  6 1365 1384 -676.3 2 vs 3   0.619  0.7339
```

So the final model would not include Group, and we can fit a common line for all therapies:

With reference to Table 7, you can show that the equation of the fitted line would be: mean ability = 30.93 + 4.76 time.

Table 7: Summary of coefficients and variance components for final mixed effects model

```
Random effects:
 Formula: ~time | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 20.128 (Intr)
time         3.040 -0.348
Residual     5.181

Fixed effects: ability ~ time
             Value Std.Error  DF t-value p-value
(Intercept) 30.930     4.190 167   7.381       0
time         4.764     0.642 167   7.426       0
```

In summary, although we can conclude that there is a significant increase in the BARTEL index with time (p$\approx$ 0) we cannot conclude that there is a significant difference amongst the three therapies. For a one week increase in time, the BARTEL index will increase by 4.76 (se =0.64), on average.

R code

```
%##### Question 1 nested model - reskill workers ##############################
% rm(list=ls())
% options(digits = 3, show.signif.stars = FALSE)
%
% # Libraries
% library(nlme)
% library(lattice)
% # read in data
% skill.df <- read.table("skill.txt", header = T)
% # declare Location and Instructor as factors and attach data set
% skill.df$Location <- factor(skill.df$Location,labels=c("C1","C2","C3"))
% skill.df$Instructor <- factor(skill.df$Instructor,
% labels=c("Instructor1","Instructor 2"))
% attach(skill.df)
%
% # exploratory plots
% bwplot(Result~Instructor|Location)
% dotplot(Instructor~Result|Location,strip=FALSE,strip.left=TRUE,layout=c(1,3),
% cex=1.5,ylab="instructors within centres",
% xlab="results",jitter.y=TRUE)
%
% ### fit 3 models and compare using AIC
% ## compare 3 models using AIC
%
% Cand.mod<-list()
% # nested model
% Cand.mod[[1]]<- lme(Result ~1 , random = ~1 | Location/Instructor)
% #random intercepts for the Location
% Cand.mod[[2]]<- lme(Result ~1 , random = ~1 | Location)
% # random intercepts for instructor
% Cand.mod[[3]]<-  lme(Result ~1 , random = ~1 | Instructor)
%
% Modnames<-c("model1","model2","model3")
%
% aict<-aictab(Cand.mod,Modnames, sort=T)
% aict
%
% ## final model --> model 2- only random effect as location
%
% summary(Cand.mod[[2]])
% intervals(Cand.mod[[2]])
%
```

```
############################ Question - stroke  ###################################
rm(list=ls())
options(digits=4, show.signif.stars=F)
library(tidyr)
library(nlme)
library(lattice)
library(ggplot2)

data.df <- read.table("stroke.txt", header =T)

#convert data to long format.
vnames <- names(data.df)
stroke <- reshape(data.df, varying = list(vnames[3:10]),
                  timevar = "Week", times = c(1:8),
                  v.names = "Ability", idvar = "Subject",direction ="long")

stroke$Group <- factor(stroke$Group)
stroke$Subject <- factor(stroke$Subject)

# Exploratory plots
stroke2 <- groupedData(Ability ~ Week|Subject, data=stroke, outer= ~Group)
# trace plot per subject
Ability.plot <- plot(stroke2, outer=T, aspect = 3)
print(Ability.plot)

# interaction plot
with(stroke2,interaction.plot(Week, Group,Ability, lwd=2))

# ggplot2 option
ggplot(stroke2, aes(Week, Ability))+
stat_smooth(aes(group = 1), method = "lm", se = T)+
stat_summary(aes(group = 1), fun = mean, geom = "point",
shape = 16, size = 1)+ facet_wrap(~Group)

#linear response for each subject
List <- lmList(Ability~Week, data=stroke2)
plot1 <- intervals(List)
plot(intervals(List))
```

```
### Fit a lme model with Random intercepts only

mod.lme <- lme(Ability ~ Group*Week, random= ~ 1 | Subject, data=stroke2)
anova(mod.lme)
summary(mod.lme)
res<-mod.lme$residuals[,2]
acf(res)

### Fit a lme model with Random slopes and random intercepts

mod2.lme <- lme(Ability ~ Group*Week, random = ~ Week|Subject, data=stroke2)
anova(mod2.lme)
summary(mod2.lme)

# Compare models
anova(mod.lme,mod2.lme)

## (g) plot the final model

with(stroke2, plot(Ability~Week, type="n"))
with(stroke2,points(Ability[Group == "A"]~Week[Group == "A"], pch=1,col="red"))
with(stroke2,points(Ability[Group == "B"]~Week[Group == "B"],pch=2, col = "blue"))
with(stroke2,points(Ability[Group == "C"]~Week[Group == "C"],pch=3, col = "black"))
abline(lm(Ability~Week, data = subset(stroke2, Group == "A")), lty=1,col ="red")
abline(lm(Ability~Week, data = subset(stroke2, Group == "B")), lty=2, col ="blue")
abline(lm(Ability~Week, data = subset(stroke2, Group == "C")), lty=3,col ="black")
legend(1,90,c("A","B","C"),lty=1:3, pch=1:3,col=c("red","blue","black"))



## refit the models remove interaction
mod2.lme <- lme(Ability ~ Group*Week,
random = ~ Week|Subject, data=stroke2, method="ML")
mod3.lme <- lme(Ability ~ Group+Week,
random = ~ Week|Subject, data=stroke2, method="ML")
mod4.lme <- lme(Ability ~ Week,
random = ~ Week|Subject, data=stroke2, method="ML")

anova(mod2.lme,mod3.lme,mod4.lme)
summary(mod4.lme) ## Final model
```