Data: Beat The Blues and Phosphate

Thomas Pattara

23rd October 2018
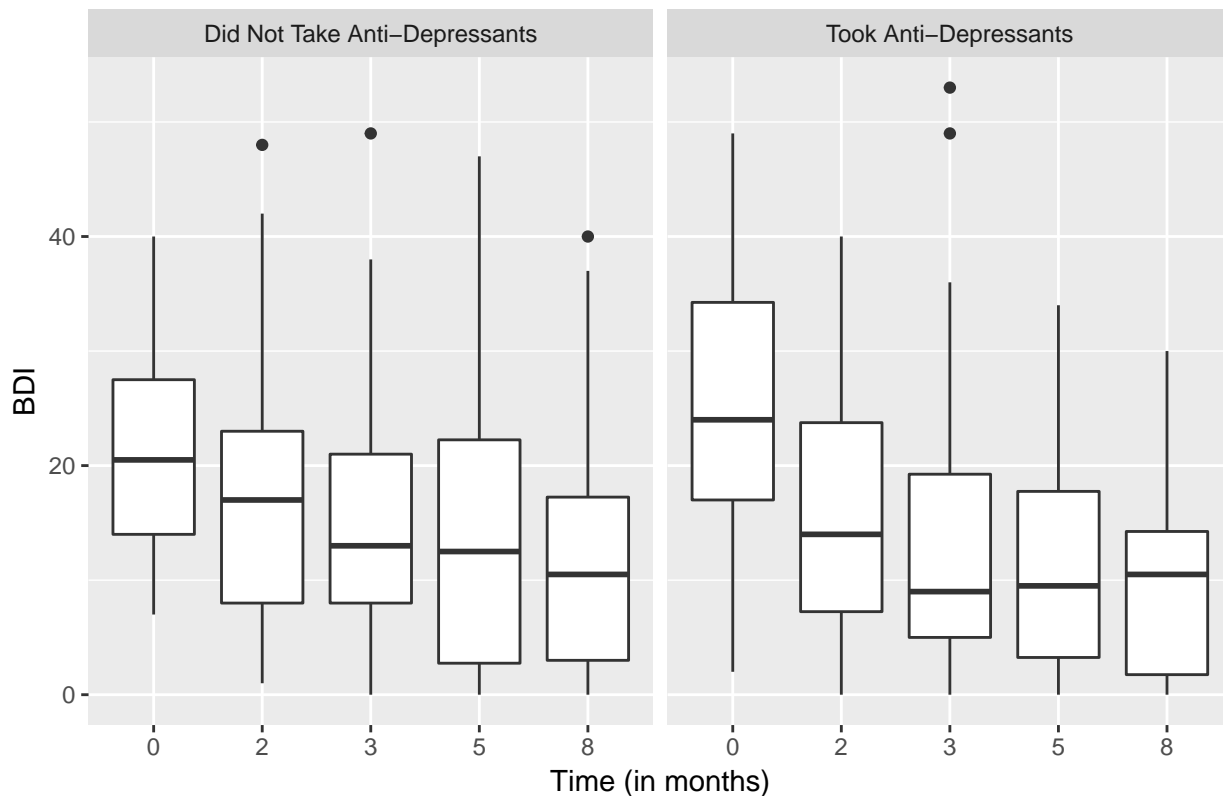
```r
library(HSAUR3)
library(dplyr)
library(tidyr)
library(ggplot2)
library(gee)
library(Matrix)
library(lme4)
library(multcomp)
library(Hmisc)
```
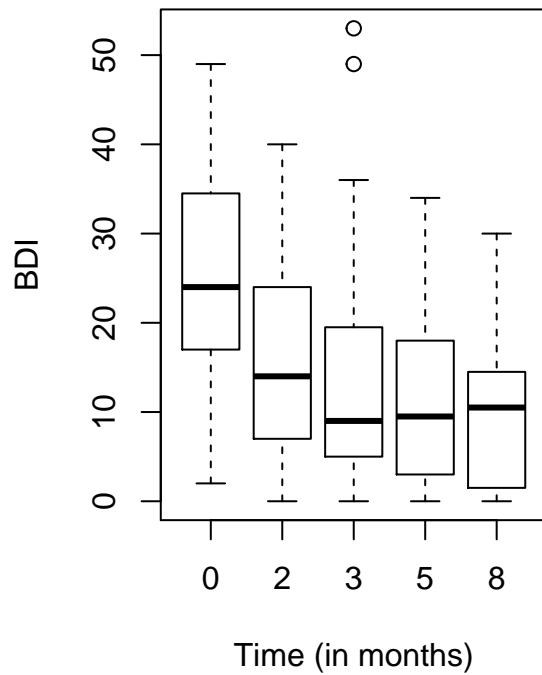
Question 1

Following up with the Beat the Blues data from the video (package HSAUR3) do the following

a. Construct boxplots to compare the factor variable drug in an analogous way to how we constructed boxplots in the video for the treatment variable. Discuss the results.**
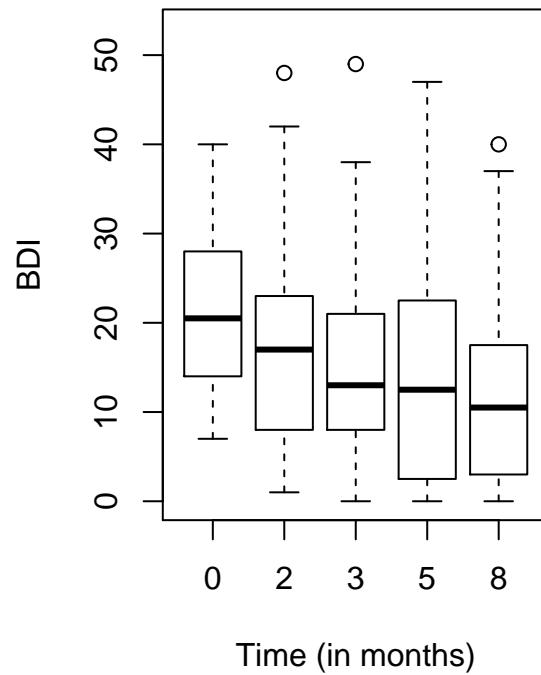
Took Anti–Depressants


**Did Not Take Anti–Depressants**

Answer:

We can see from the box plot that the Beck Depression Inventory II (BDI) decreases with lower variability in patients who took anti-depressants. The slight decrease in BDI in patients who did not take anti-depressants and the high variability suggests that patients are more likely to reducte their BDI levels with anti-depressant treatment as time increases.

b. Repeat (a) for the length variable. Discuss the results.

GGPLOT Length of Current Episode of Depression
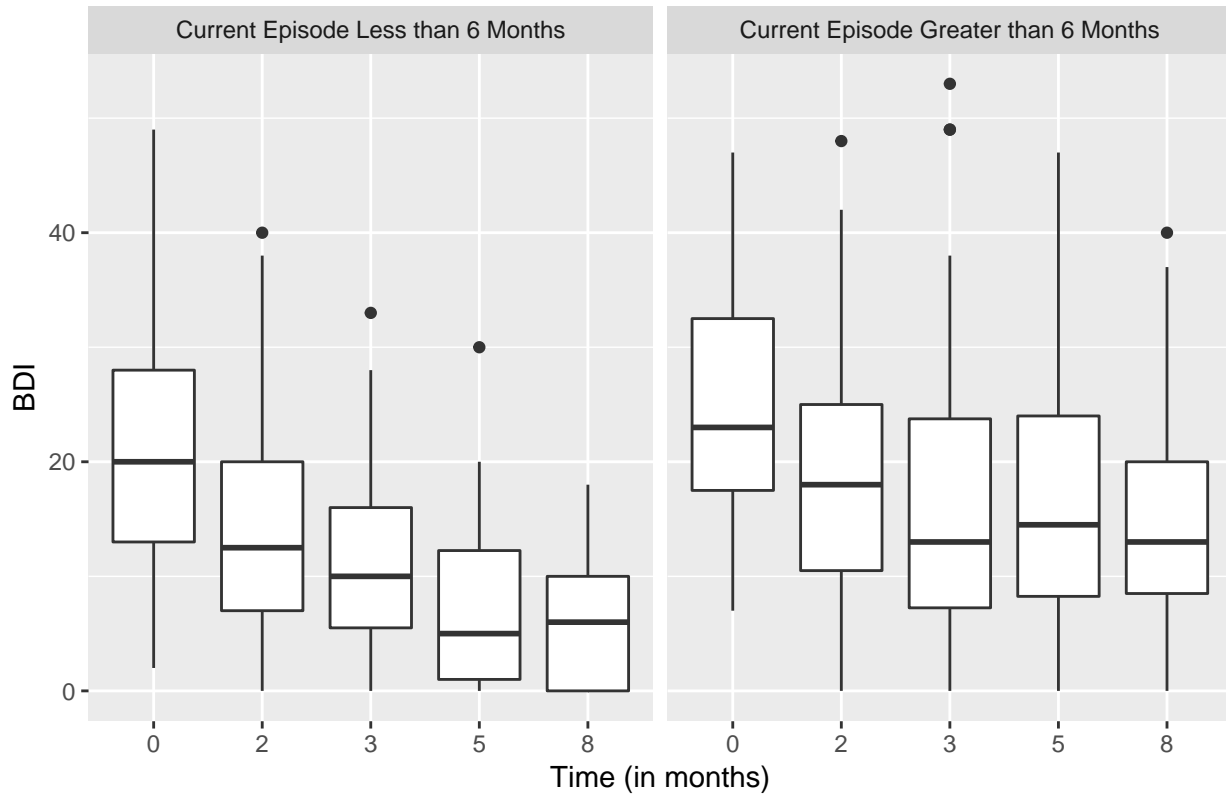
3
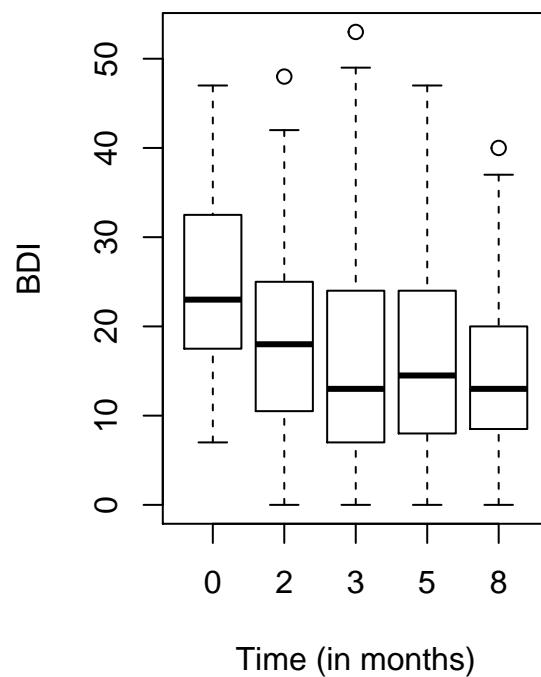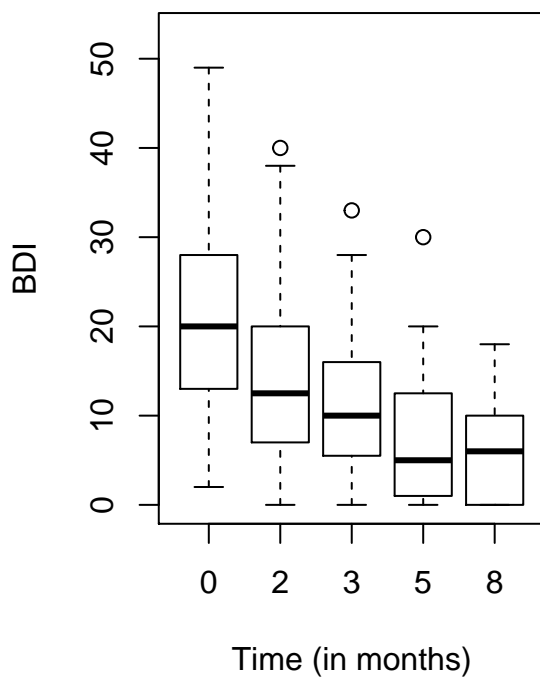
Answer:

We can clearly see that the Beck Depression Inventory II (BDI) levels decrease greatly in patients that have shorter episodes of depression than that of patients that have longer episodes of depression. There is much more variablility in BDI levels for patients who have longer episodes of depression.

c. Use the *lm* function to fit a model to the Beat the Blues data that assumes that the repeated measurements are independent. Compare the results to those from fitting the random intercept model *BtheB_lmer1* from the video.

```
## Data: btb
## Models:
## BtheB_lmer1: bdi ~ bdi.pre + time + treatment + drug + length + (1 | subject)
## BtheB_lm1: bdi ~ bdi.pre + time + treatment + drug + length + subject
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## BtheB_lmer1   8 1887.5 1916.6 -935.75   1871.5
## BtheB_lm1    99 1778.2 2138.0 -790.08   1580.2 291.32     91  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = bdi ~ bdi.pre + time + treatment + drug + length +
##     (1 | subject), data = btb, REML = FALSE, na.action = na.omit)
##
## Linear Hypotheses:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0    5.59239    2.24244   2.494   0.0126 *
## bdi.pre == 0        0.63968    0.07789   8.212 2.22e-16 ***
## time == 0          -0.70476    0.14639  -4.814 1.48e-06 ***
## treatmentBtheB == 0 -2.32908    1.67036  -1.394   0.1632
## drugYes == 0       -2.82495    1.72684  -1.636   0.1019
## length>6m == 0      0.19708    1.63832   0.120   0.9043
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

Random intercepts

Residuals

Normal Q–Q

lm(bdi ~ bdi.pre + time + treatment + drug + length + subject)
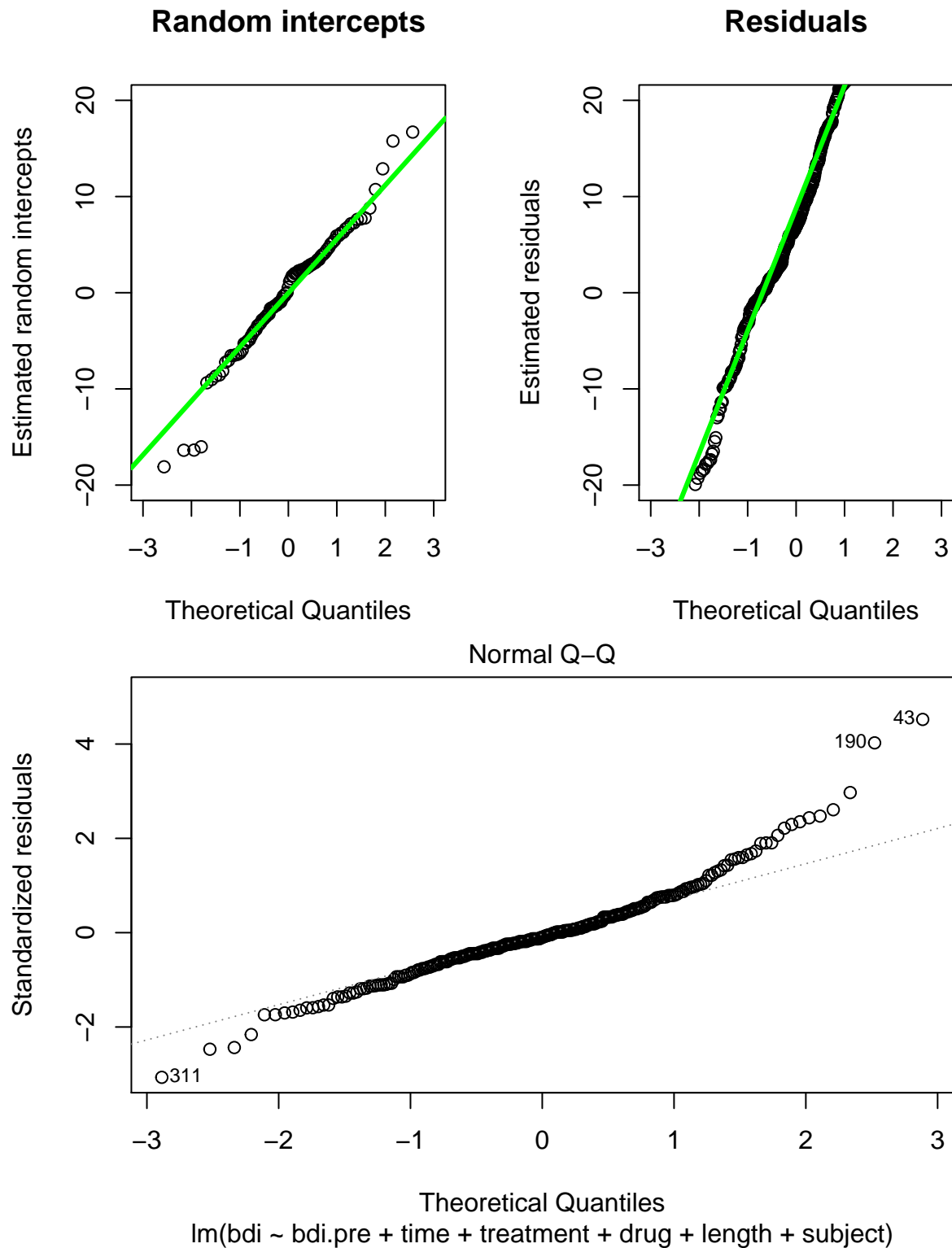
Answer:

The linear model produces a lower AIC than the linear mixed-effect model. The residuals plot of the mixed-effects model vs the linear model is slightly misleading, but the Q-Q plot suggests the residuals are

normally distributed. The significance of chi-squared test and the lower AIC suggests that the linear model is an adequate model.

d. Investigate and discuss whether there is any evidence of an interaction between treatment and time for the Beat the Blues data.
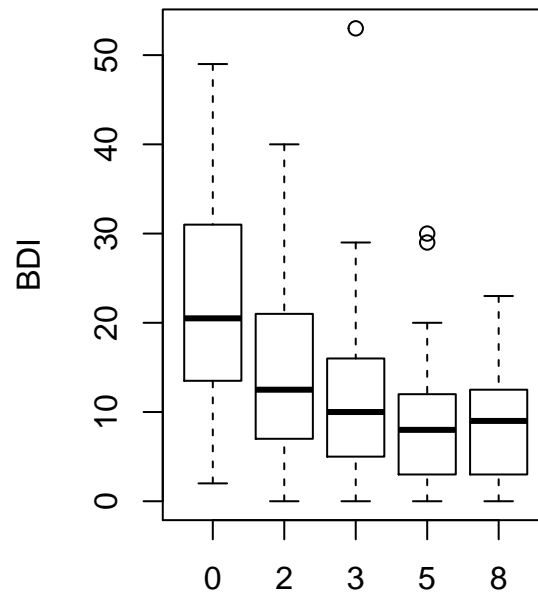

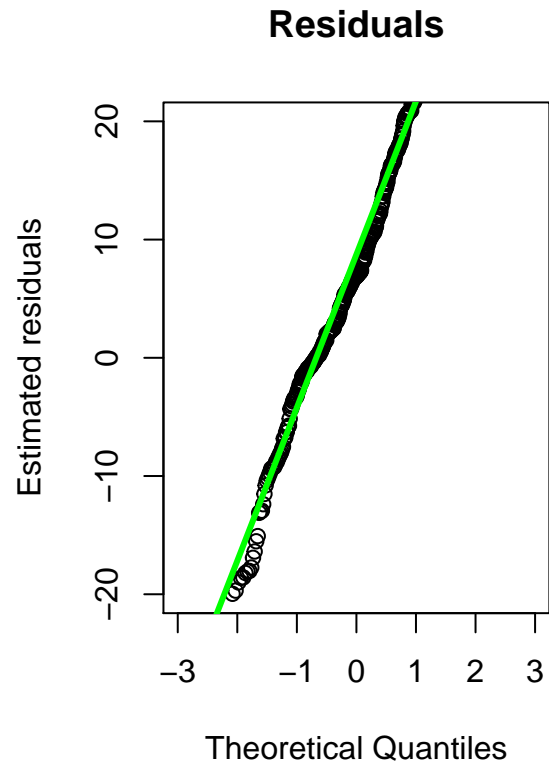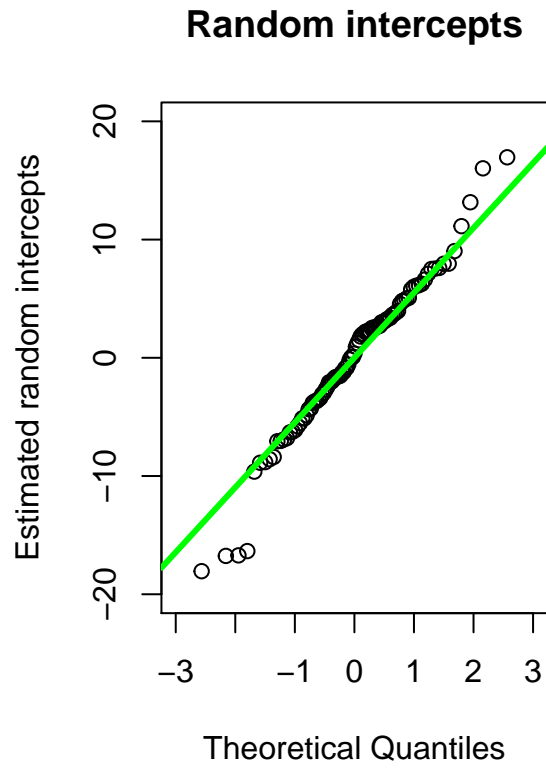GGPLOT Treatment Type
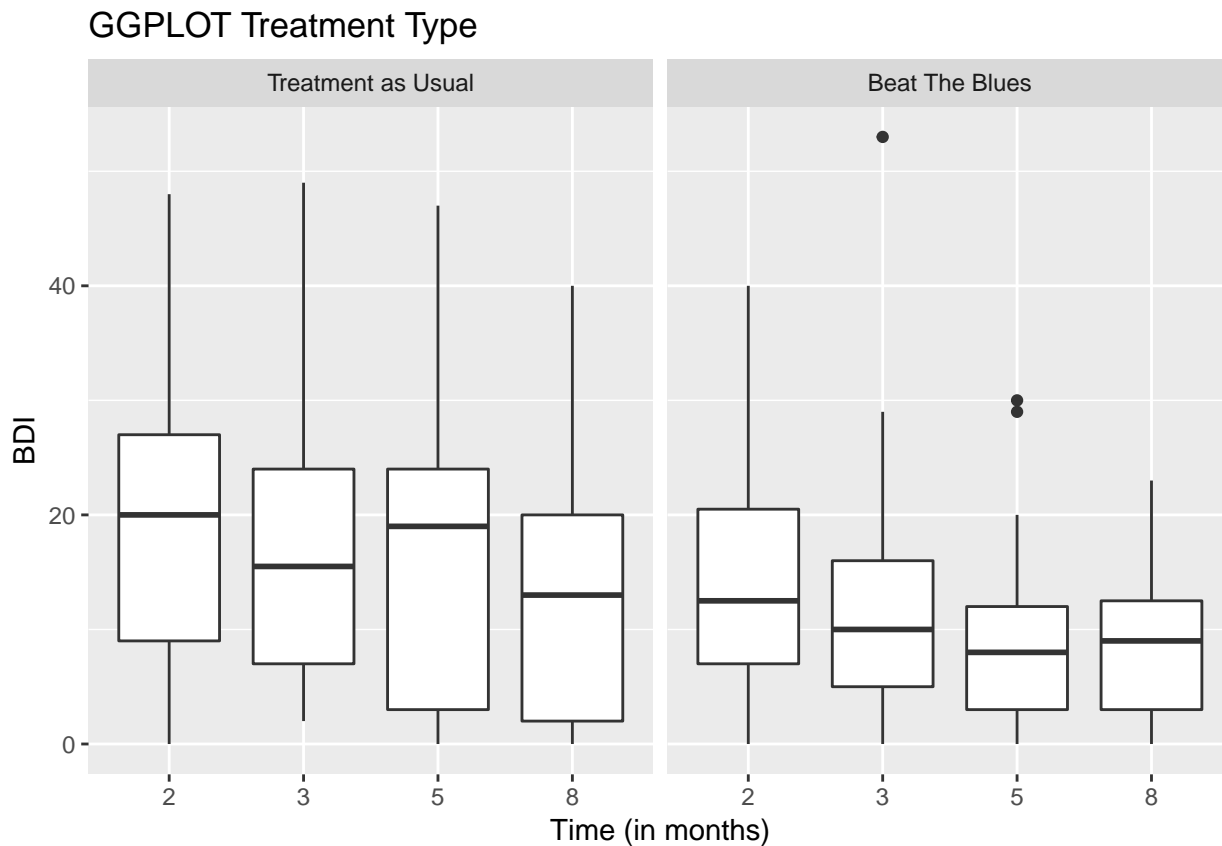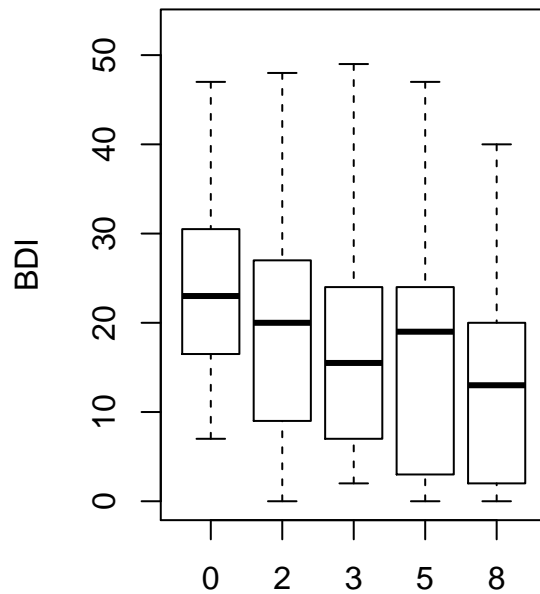
```
##                          Estimate Std. Error  t value   Pr(>|t|)
## time:treatmentBtheB 0.6435779  0.2975651 2.162814 0.03186724

##
##    Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = bdi ~ bdi.pre + time + treatment + drug + length +
##      time * treatment + (1 | subject), data = btb, REML = FALSE,
##      na.action = na.omit)
##
## Linear Hypotheses:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0          6.47824    2.31905    2.793  0.00521 **
## bdi.pre == 0              0.64046    0.07843    8.166 2.22e-16 ***
## time == 0                -0.95555    0.20831   -4.587 4.49e-06 ***
## treatmentBtheB == 0      -4.09804    1.98490   -2.065  0.03896 *
## drugYes == 0             -2.79209    1.73986   -1.605  0.10854
## length>6m == 0            0.21905    1.65043    0.133  0.89441
## time:treatmentBtheB == 0  0.49000    0.28959    1.692  0.09064 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

Random intercepts — Estimated random intercepts vs Theoretical Quantiles

Residuals — Estimated residuals vs Theoretical Quantiles

Answer:

The Beat the Blues treatment type displays downward trend with lower variance as time increases, hence there is an interaction between time and treatment type. This also indicates that the Beat the Blues treatment does reduce Back Depression Inventory II (BDI) as time increases.
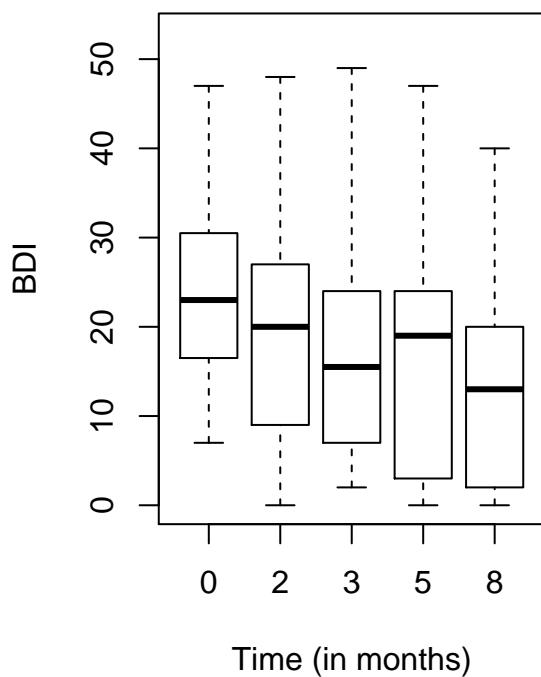
e. Construct a plot of the mean profiles of both treatment groups in the Beat the Blues data, showing also standard deviation bars at each time point.
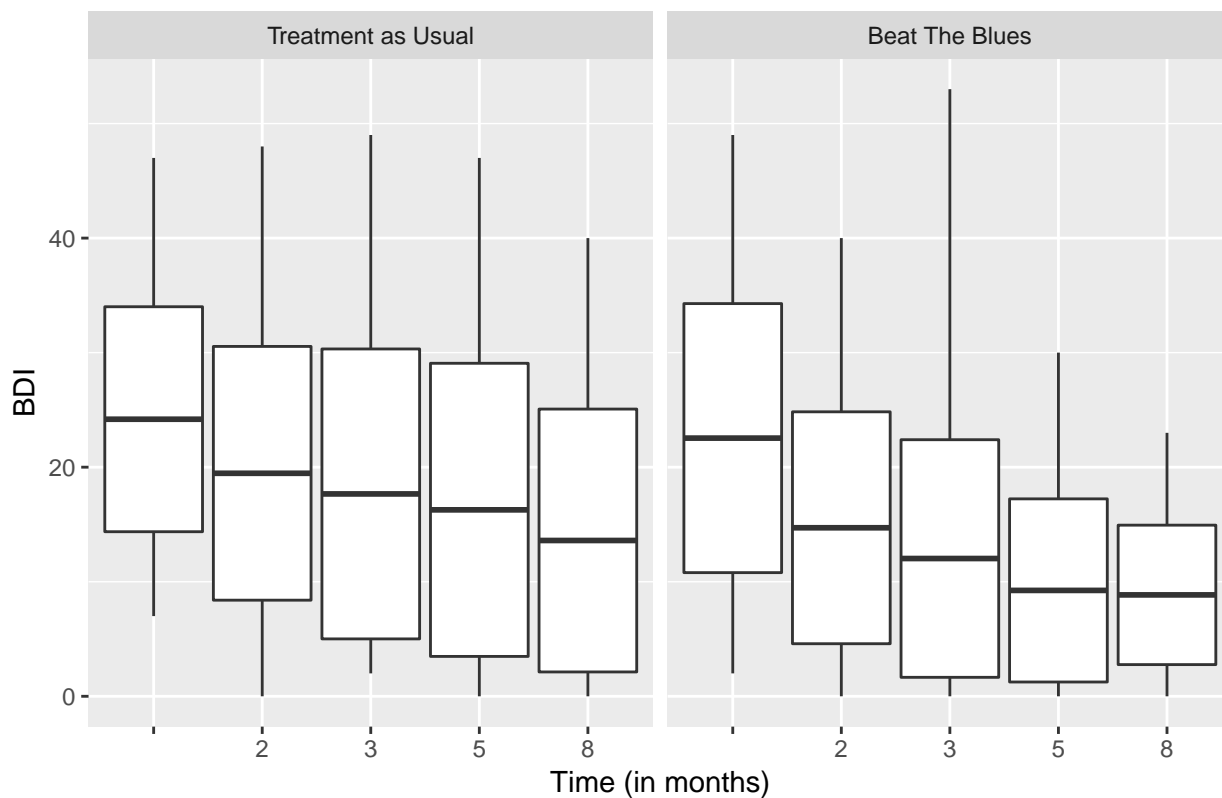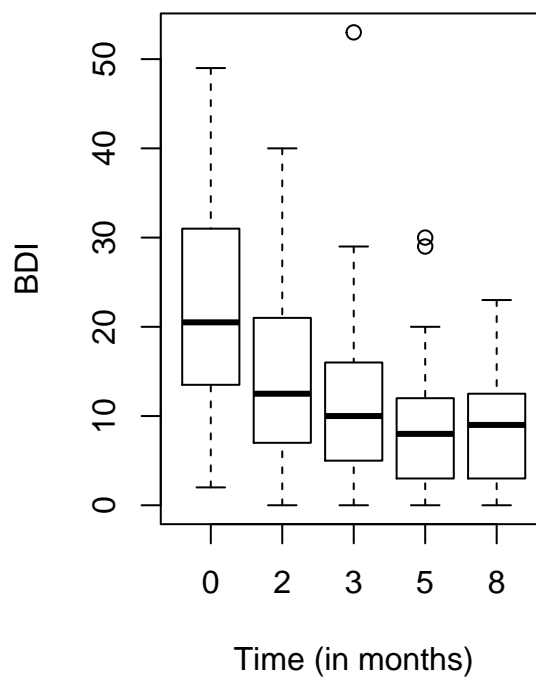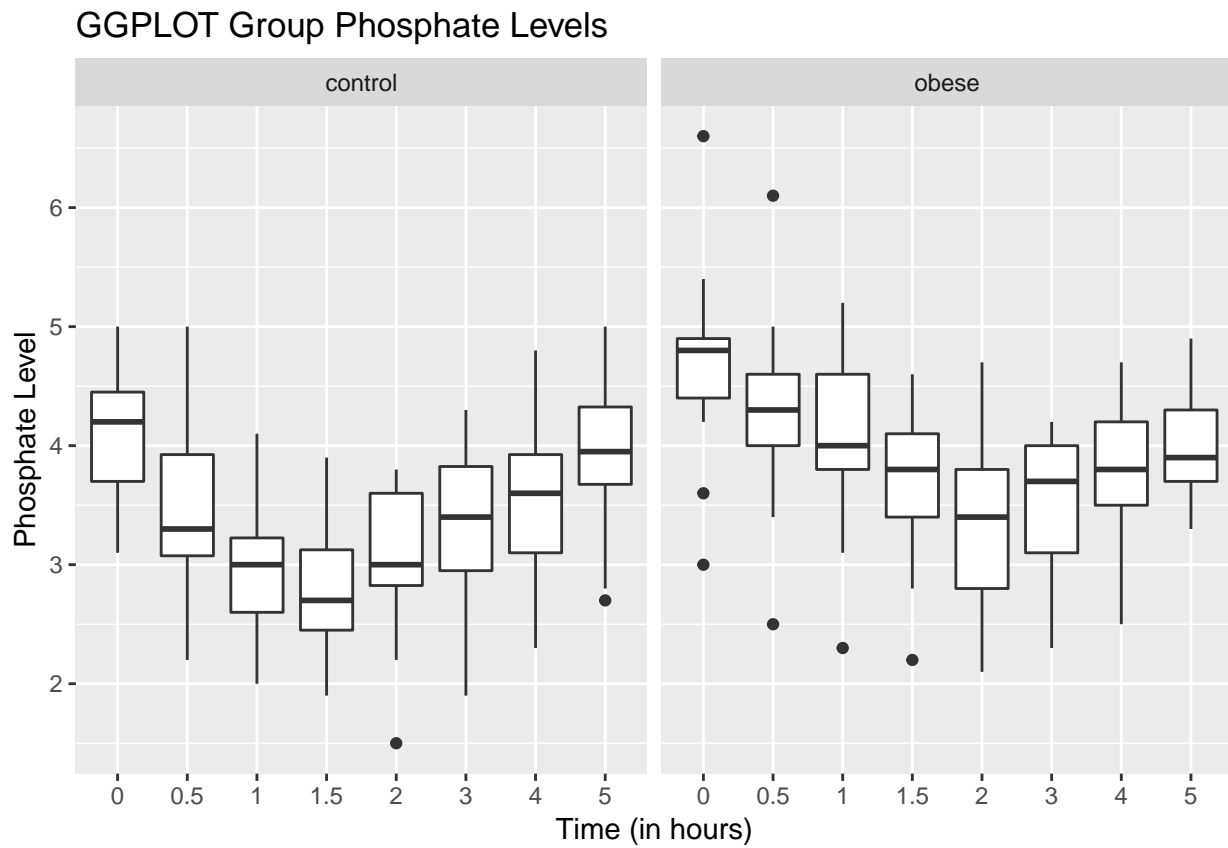
Question 2

Consider the **phosphate** data from the package HSAUR3. This data shows the plasma inorganic phosphate levels for 33 subjects, 20 of whom are controls and 13 of whom have been classified as obese (Davis, 2002). Perform the following on this dataset

a. Construct boxplots by group and discuss.
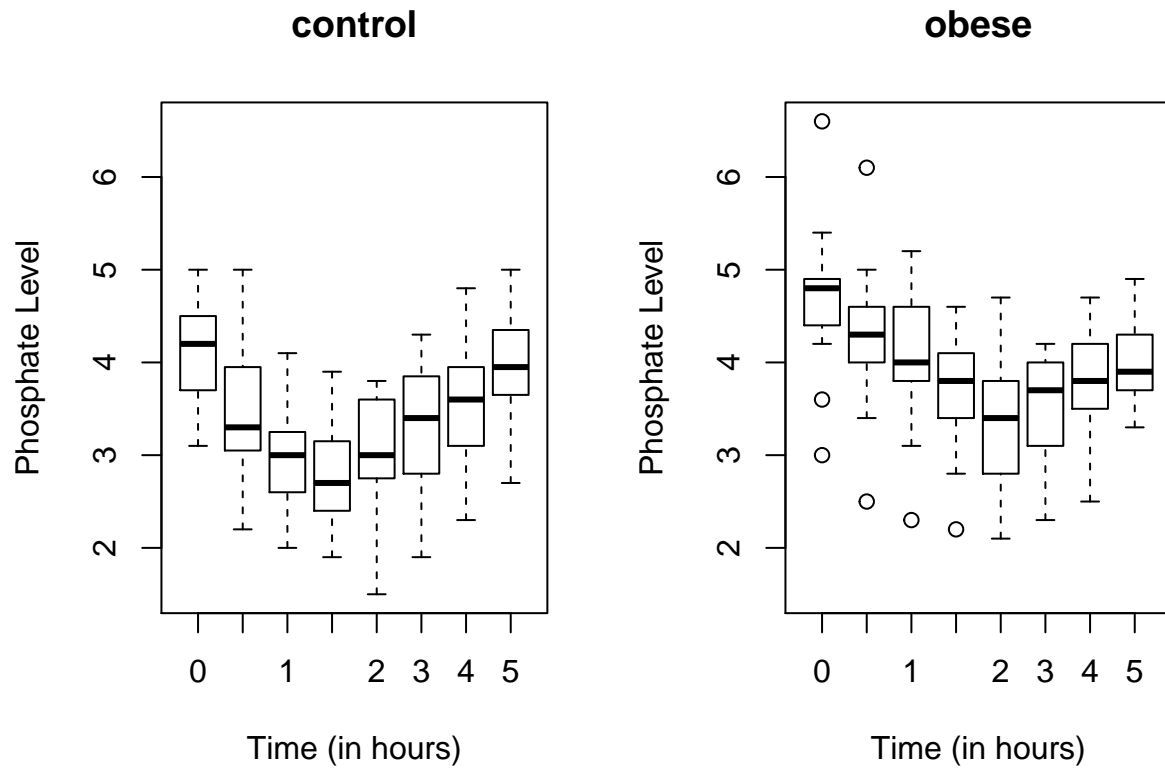
Answer:

The phosphate levels in the control group can be obseeved to drop till the first 1 and a half hours and then rise for the next following hours, compared to obese group, which tend to slightlty drop untill the second hour and then rise again for the following hours, comparatively more variability is seen for the obese group.

b. Produce separate plots of the profiles of the individuals in each group.

GGPLOT Group Phosphate Levels By Time & Subject
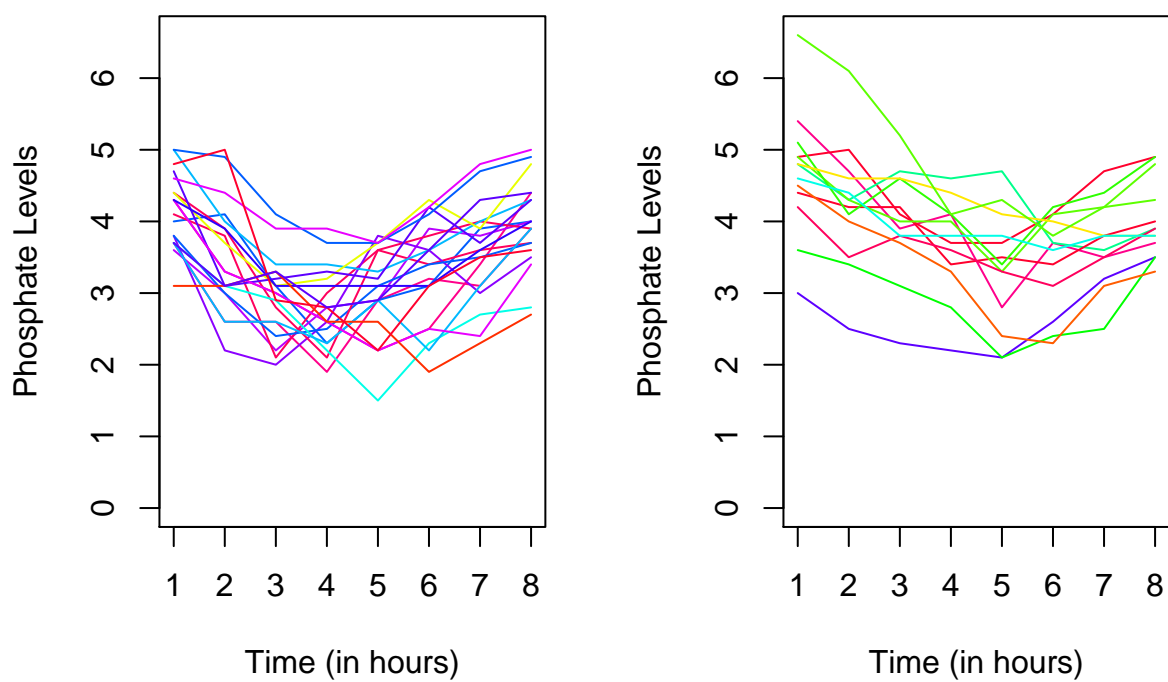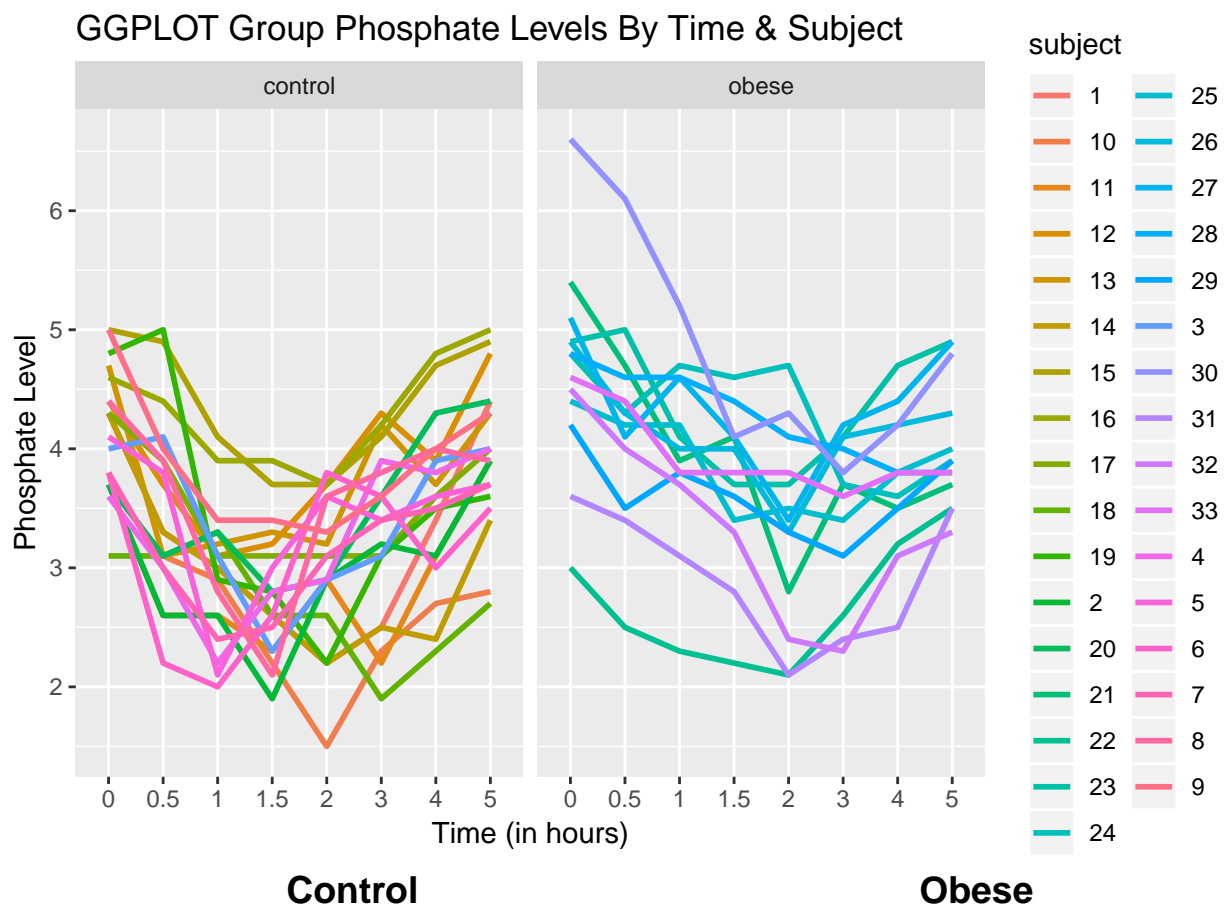
c. Guided by how these plots fit, which linear mixed effects models do you think might be sensible? (Hint: Discuss intercept and slope, intercept and interaction).

It would be sensible to fit a linear mixed effects model because the slopes and intercepts in each group for the model vary enormously. There is an interaction with respect to time within the two group variables control and obese.

**d. Convert the data to long version and fit the model of your choice and discuss the results.**

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = phosphate_level ~ t0 + time * group + (1 | subject),
##     data = phos)
##
## Linear Hypotheses:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0          0.88700    0.38907   2.280 0.022621 *
## t0 == 0                   0.63084    0.08954   7.045 1.85e-12 ***
## time1 == 0               -0.55500    0.13424  -4.134 3.56e-05 ***
## time1.5 == 0             -0.72000    0.13424  -5.364 8.16e-08 ***
## time2 == 0               -0.50500    0.13424  -3.762 0.000169 ***
## time3 == 0               -0.20500    0.13424  -1.527 0.126731
## time4 == 0                0.06000    0.13424   0.447 0.654903
## time5 == 0                0.45500    0.13424   3.389 0.000700 ***
## groupobese == 0           0.40106    0.18965   2.115 0.034458 *
## time1:groupobese == 0     0.31654    0.21388   1.480 0.138874
## time1.5:groupobese == 0   0.18154    0.21388   0.849 0.395996
## time2:groupobese == 0    -0.38731    0.21388  -1.811 0.070159 .
## time3:groupobese == 0    -0.57192    0.21388  -2.674 0.007494 **
## time4:groupobese == 0    -0.58308    0.21388  -2.726 0.006407 **
## time5:groupobese == 0    -0.67038    0.21388  -3.134 0.001722 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)

## Linear mixed model fit by REML ['lmerMod']
## Formula: phosphate_level ~ t0 + time * group + (1 | subject)
##    Data: phos
##
## REML criterion at convergence: 330.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3420 -0.6072  0.0432  0.6218  3.2895
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  subject   (Intercept) 0.08565  0.2927
##  Residual              0.18020  0.4245
## Number of obs: 231, groups:  subject, 33
##
## Fixed effects:
##                   Estimate Std. Error t value
```

```
## (Intercept)          0.88700    0.38907    2.280
## t0                   0.63084    0.08954    7.045
## time1               -0.55500    0.13424   -4.134
## time1.5             -0.72000    0.13424   -5.364
## time2               -0.50500    0.13424   -3.762
## time3               -0.20500    0.13424   -1.527
## time4                0.06000    0.13424    0.447
## time5                0.45500    0.13424    3.389
## groupobese           0.40106    0.18965    2.115
## time1:groupobese     0.31654    0.21388    1.480
## time1.5:groupobese   0.18154    0.21388    0.849
## time2:groupobese    -0.38731    0.21388   -1.811
## time3:groupobese    -0.57192    0.21388   -2.674
## time4:groupobese    -0.58308    0.21388   -2.726
## time5:groupobese    -0.67038    0.21388   -3.134
```
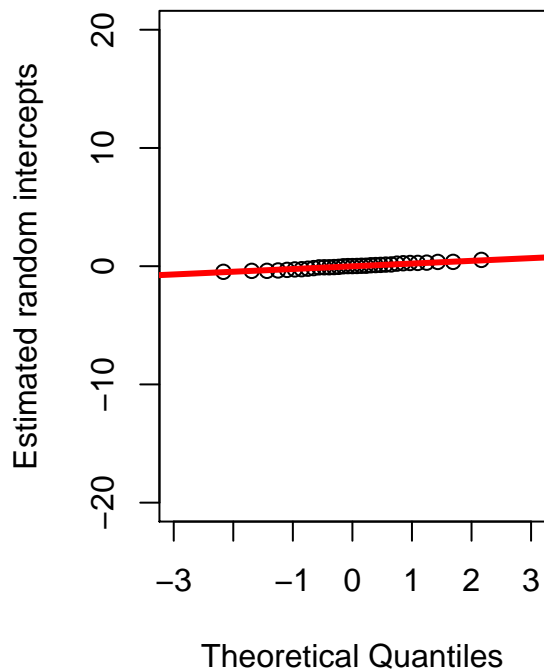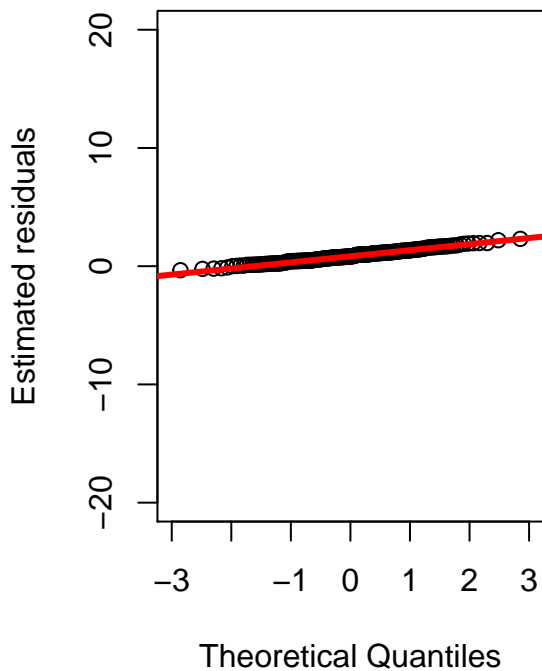


```
## [1] 0.151
```

Answer:

Looking at the coefficients of the model, there is a significant interaction between the group and time, as noted earlier by the box plot. The random intercepts and residuals of the model seem to be normally distributed with center around zero. The mean squared error of the fixed-effects model is 0.151 which suggests that the model is good and adequate.