Problem 1.

We consider the model: $Y_{ij} = \beta_0 + \beta_1 X_{ij} + b_{0i} + b_{1i} W_{1,ij} + b_{2i} W_{2,ij} + \varepsilon_{ij}$,

$$j = 1(1)n_i, \quad i = 1(1)n$$

Also, $\underset{\sim}{b_i} \sim N(\underset{\sim}{0}, D)$, where, $D = \begin{bmatrix} 4 & 1 & 0.5 \\ 1 & 2 & 0.5 \\ 0.5 & 0.5 & 2 \end{bmatrix}$

The matrix notation of the model is: $\underset{\sim}{Y_i} = X\underset{\sim}{\beta} + Z\underset{\sim}{b} + \underset{\sim}{\varepsilon_i}$

(a) Let, $n_i = 3$ and $\sigma^2 = 3$,

So, $Z = \begin{bmatrix} 1 & W_{1,i1} & W_{2,i1} \\ 1 & W_{1,i2} & W_{2,i2} \\ 1 & W_{1,i3} & W_{2,i3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}$

So, the conditional distribution of the given mixed effects model is:

$$\underset{\sim}{Y_i} \mid \underset{\sim}{b_i} \sim N\left(X\underset{\sim}{\beta} + Z\underset{\sim}{b_i}, \, \varepsilon_i \right)$$

$$\underset{\sim}{b_i} \sim N(\underset{\sim}{0}, D)$$

Also, the marginal distribution is: $\underset{\sim}{Y_i} \sim N(X\underset{\sim}{\beta}, ZDZ' + \varepsilon_i)$

where, $\varepsilon_i = \sigma^2 I_3 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

Thus, the marginal variance-covariance matrix of $\underset{\sim}{Y_i}$ is:

$$ZDZ' + \varepsilon_i = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 & 0.5 \\ 1 & 2 & 0.5 \\ 0.5 & 0.5 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 5 & 3 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 71 & 51.5 & 24 \\ 51.5 & 46 & 21.5 \\ 24 & 21.5 & 15 \end{bmatrix}$$

(b) Let, $n_i = 2$ and $\sigma^2 = 2$,

So, $Z = \begin{bmatrix} 1 & W_{1,i1} \\ 1 & W_{1,i2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

Thus, the conditional variance-covariance matrix of $Y_i | b_i$ is:

$$\Sigma_i = \sigma^2 I_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}_{\square}$$

(c) Here, we test: $H_0: b_i \in \Theta_{b,0}$ vs $H_a:$ not H_0, where $\Theta_{b,0}$ is

Thus, we are testing whether the random effect parameters are dependent.

a subset of the parameter space Θ_b of the random effects.

under H_0, $-2\log L_0 = 340$

and $-2\log L_a = 324$ (with 2 degrees of freedom).

Test statistic $= -2\log\left(\dfrac{L_0}{L_a}\right) = \cancel{-2\log(340 - 324)} \; 340 - 324$

$= 16$

Also, under H_0, the test statistic $\sim \chi^2_{df}$

Problem 2

The model reads as: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + e_{ij}$; $i = 1(1)8$ — (1)

where, Y_{ij} = average leaf weight per plant on plot i at time point t_{ij}.

If we add a random effect b_{1i}, the model becomes:

$$Y_{ij} = \beta_0 + (\beta_1 + b_{1i}) t_{ij} + \beta_2 t_{ij}^2 + e_{ij}, \quad — (2)$$

where, $b_{1i} \overset{iid}{\sim} N(0, \sigma_b^2)$, $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

(a) I feel that on graphing the average leaf weights for each of the eight plots, the lines might differ from each other for the difference in slopes. For this purpose, adding a random component to the fixed effect model would be better in making a cluster and compare the trends for the different plots.

(b) The matrix form of model (2) is:

$$Y_i = X_i \beta + z_i b_i + e_i$$

$$Y = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{8j} \end{bmatrix}^{8 \times 1}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \end{bmatrix}, \quad X_i = \begin{bmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i}^2 \end{bmatrix}^{n_i \times 3}$$

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{bmatrix}^{n_i \times 1}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}^{3 \times 1}, \quad z_i = \begin{bmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{in_i} \end{bmatrix}^{n_i \times 1},$$

$$b_i = \begin{bmatrix} b_{1i} \end{bmatrix}^{1 \times 1}, \quad var(b_i) = \sigma_b^2$$

(c) The marginal variance - covariance matrix of Y_i is:

$$\sigma_b^2 Z_i Z_i' + \sigma^2 I_{ni} = \sigma_b^2 \begin{bmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{ini} \end{bmatrix} \begin{bmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{ini} \end{bmatrix}' + \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma_b^2 \begin{bmatrix} t_{i1}^2 & t_{i1} t_{i2} & \cdots & t_{i1} t_{ini} \\ t_{i2} t_{i1} & t_{i2}^2 & \cdots & t_{i2} t_{ini} \\ \vdots & & & \\ t_{ini} t_{i1} & t_{ini} t_{i2} & \cdots & t_{ini}^2 \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

$$= \sigma_b^2 \begin{bmatrix} t_{i1}^2 + \frac{\sigma^2}{\sigma_b^2} & t_{i1} t_{i2} & \cdots & t_{i1} t_{ini} \\ t_{i2} t_{i1} & t_{i2}^2 + \frac{\sigma^2}{\sigma_b^2} & \cdots & t_{i2} t_{ini} \\ \vdots & & & \vdots \\ t_{ini} t_{i1} & t_{ini} t_{i2} & \cdots & t_{ini}^2 + \frac{\sigma^2}{\sigma_b^2} \end{bmatrix} \quad n_i \times n_i$$

∎

(d) Let, $\sigma^2 = \sigma_b^2 = 1$, $t_{i1} = 1$, $t_{i2} = 2$, $t_{i3} = 3$

Now,

$$cor(Y_{ij}, Y_{ij'}) = \frac{\sigma_b^2 \, t_{ij} \, t_{ij'}}{\sqrt{\sigma_b^2 \left(t_{ij}^2 + \frac{\sigma^2}{\sigma_b^2}\right)} \cdot \sqrt{\sigma_b^2 \left(t_{ij'} + \frac{\sigma^2}{\sigma_b^2}\right)}}$$

considering $j=1$ and $j'=2$,

$$cor(Y_{i1}, Y_{i2}) = \frac{t_{i1} t_{i2}}{\sqrt{(t_{i1}^2 + 1)} \sqrt{(t_{i2}^2 + 1)}} = \frac{1 \times 2}{\sqrt{2} \sqrt{5}} = \frac{2}{\sqrt{10}} = 0.63$$

considering $j=2$ and $j'=3$, $cor(Y_{i2}, Y_{i3}) = \frac{t_{i2} t_{i3}}{\sqrt{(t_{i2}^2 + 1)} \sqrt{(t_{i3}^2 + 1)}} = \frac{3}{\sqrt{20}} = 0.67$

So the correlations are not decreasing with temporal separation.

I don't feel that the results are realistic because with temporal separation, the correlation should decrease when accounting for several other factors, including environmental changes.

Part (e)

Hierarchically modeling the marginal covariance as in model 2 allows for the incorporation of plot-specific random effects, capturing individual variability in growth patterns while also accommodating the overall variation in growth trajectories over time. This approach provides a more flexible and interpretable framework for modeling the complex covariance structure observed in longitudinal data. In contrast, directly modeling the marginal covariance matrix without hierarchical structuring may lack the ability to capture the inherent variability in growth patterns among different plots, potentially leading to oversimplified or inaccurate representations of the data. Hierarchical modeling allows for a more nuanced understanding of the underlying processes driving the observed covariance structure, thereby enhancing the robustness and interpretability of the statistical analysis.

PROBLEM 3.

```r
library(dplyr)
library(lmerTest)
library(MCMCglmm)
library(pbkrtest)
library(RLRsim)
library(agridat)
library(latticeExtra)
library(lme4)
library(nlme)
library(faraway)
library(ez)
library(ggplot2)
library(easyanova)
library(cowplot)

# Load the data
data(data8)
```

Part (a)

```r
# Plot milk yield by block, with pasture and mineral factor levels
library(ggplot2)
ggplot(data8, aes(x = block, y = milk, shape = pasture,
                  color = mineral)) + geom_point(size = 2)
```
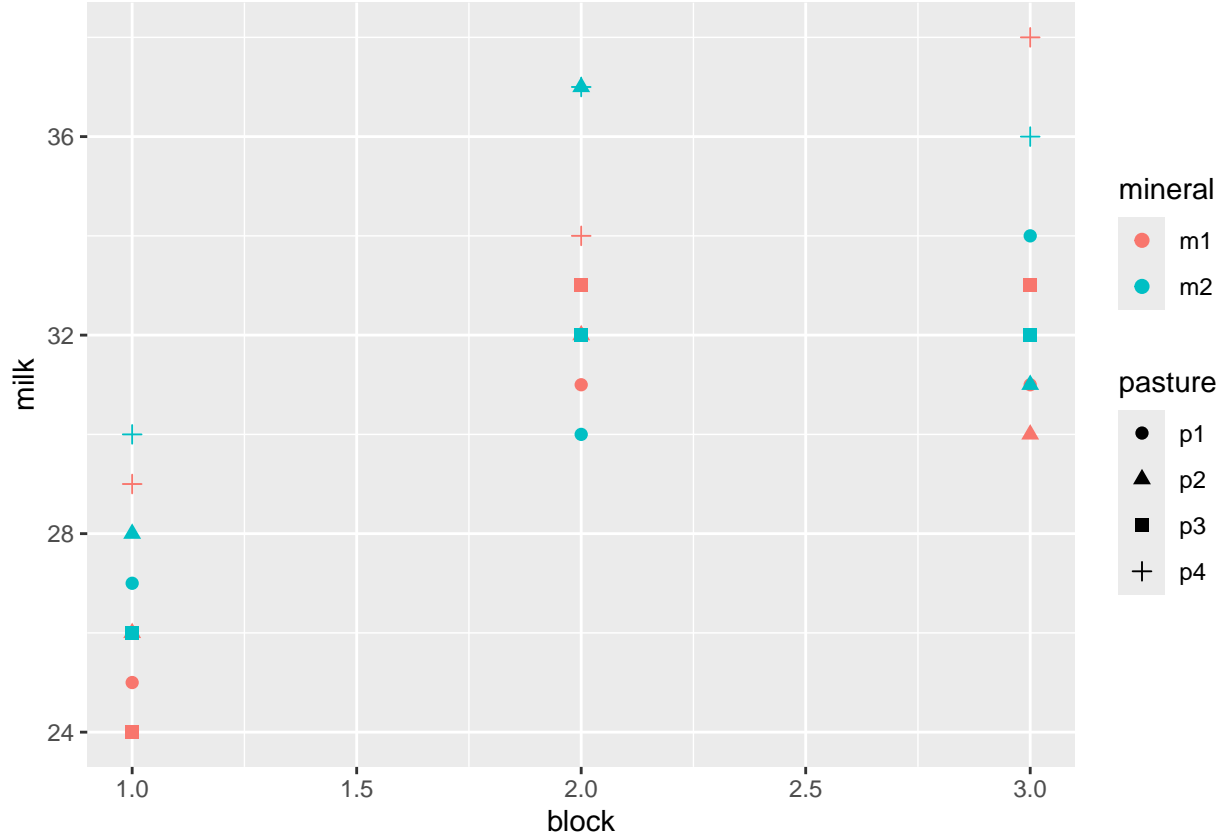
Interpretation : The given plot helps to visualize any patterns or trends in milk yield across blocks and how they might be influenced by pasture and mineral factors. It mainly depicts the variability in milk yield across different blocks and also within a single block on the basis of different level combinations of minerals supplements and pasture treatments.

Part (b)

The model used for the given split-plot design is :

$Y_{ijk} = \mu + \rho_i + \alpha_j + (\rho\alpha)_{ij} + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$, i = 1(|)3, j = 1(|)4 & k = 1(|)2, $\epsilon_{ijk}{\sim}N(0, \sigma^2)$

where, Y_{ijk} is the response corresponding to the ith block, jth pasture and kth mineral supplement

μ is the overall mean.

ρ_i is the effect of the random blocking.

α_j is the fixed effect of the ith pasture.

β_k is the fixed effect of the kth mineral supplement.

$(\rho\alpha)_{ij}$ is the random interaction effect of the ith block and jth pasture.

$(\alpha\beta)_{jk}$ is the interaction effect of the jth pasture and the kth mineral supplement.

ϵ_{ijk} is the random subplot error.

But, there is a problem with the given model that the random effects group the data and examine effects within those groupings. Within each block, each pasture is present only one time. So, there is no way to capture variation for block:pasture interactions because they are only observed once. Thus, we need to remove the $(\rho\alpha)_{ij}$ term from the model. Thus, the final model turns out to be :

$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$, i = 1(|)3, j = 1(|)4 & k = 1(|)2, $\epsilon_{ijk}{\sim}N(0, \sigma^2)$

```
mixed_model <- lmer(milk ~ pasture*mineral-1 + (1 | block), data = data8)
summary(mixed_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: milk ~ pasture * mineral - 1 + (1 | block)
##    Data: data8
##
## REML criterion at convergence: 79.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.68002 -0.21507  0.02888  0.42284  1.63288
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  block    (Intercept) 12.893   3.591
##  Residual              3.149   1.774
## Number of obs: 24, groups:  block, 3
##
## Fixed effects:
##                    Estimate Std. Error      df t value Pr(>|t|)
## pasturep1           29.0000     2.3124  2.8977  12.541 0.001294 **
## pasturep2           29.3333     2.3124  2.8977  12.685 0.001252 **
## pasturep3           30.0000     2.3124  2.8977  12.973 0.001174 **
## pasturep4           33.6667     2.3124  2.8977  14.559 0.000844 ***
## mineralm2            1.3333     1.4489 14.0000   0.920 0.373022
## pasturep2:mineralm2  1.3333     2.0490 14.0000   0.651 0.525762
## pasturep3:mineralm2 -1.3333     2.0490 14.0000  -0.651 0.525762
## pasturep4:mineralm2 -0.6667     2.0490 14.0000  -0.325 0.749719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            pstrp1 pstrp2 pstrp3 pstrp4 mnrlm2 pst2:2 pst3:2
## pasturep2    0.804
## pasturep3    0.804  0.804
## pasturep4    0.804  0.804  0.804
## mineralm2   -0.313  0.000  0.000  0.000
## pstrp2:mnr2  0.222 -0.222  0.000  0.000 -0.707
## pstrp3:mnr2  0.222  0.000 -0.222  0.000 -0.707  0.500
## pstrp4:mnr2  0.222  0.000  0.000 -0.222 -0.707  0.500  0.500
```

```
anova(mixed_model)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                Sum Sq Mean Sq NumDF  DenDF F value   Pr(>F)
## pasture        713.53 178.382     4  2.466 56.6506 0.008314 **
## mineral          8.17   8.167     1 14.000  2.5936 0.129607
## pasture:mineral  5.83   1.944     3 14.000  0.6175 0.615036
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
exactRLRT(mixed_model)
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 19.012, p-value < 2.2e-16
```

Interpretation : From the ANOVA table, it is evident that the pasture is a highly significant covariate which can be used in studying the milk yield across different blocks, but, mineral supplements and the interaction between pastures and mineral supplements are not significant. Also from the lmer() summary, we can see that all four of the pasture treatments are also significant. Also from the random LR test, we reject the null hypothesis at 5% level of significance and conclude that the variance of the random block effect is greater than 0.

Part (c)

The model used for the given split-plot design is :

$$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}, \text{ i} = 1(|)3, \text{ j} = 1(|)4 \text{ \& k} = 1(|)2, \epsilon_{ijk} \sim N(0, \sigma^2)$$

where, Y_{ijk} is the response corresponding to the ith block, jth pasture and kth mineral supplement

μ is the overall mean.

ρ_i is the effect of the random blocking.

α_j is the fixed effect of the ith pasture.

β_k is the fixed effect of the kth mineral supplement.

ϵ_{ijk} is the random subplot error.

```
mixed_model_no_inte <- lmer(milk ~ pasture+mineral-1 + (1 | block), data = data8)
summary(mixed_model_no_inte)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: milk ~ pasture + mineral - 1 + (1 | block)
##    Data: data8
##
## REML criterion at convergence: 90.5
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -1.3754 -0.3496 -0.1110  0.4914  2.1261
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  block    (Intercept) 12.919   3.594
##  Residual              2.936   1.714
## Number of obs: 24, groups:  block, 3
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
```

4

```
## pasturep1  29.0833      2.2177  2.4629  13.114  0.00243 **
## pasturep2  30.0833      2.2177  2.4629  13.565  0.00224 **
## pasturep3  29.4167      2.2177  2.4629  13.264  0.00237 **
## pasturep4  33.4167      2.2177  2.4629  15.068  0.00174 **
## mineralm2   1.1667      0.6996 17.0000   1.668  0.11368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           pstrp1 pstrp2 pstrp3 pstrp4
## pasturep2  0.900
## pasturep3  0.900  0.900
## pasturep4  0.900  0.900  0.900
## mineralm2 -0.158 -0.158 -0.158 -0.158
```

```
anova(mixed_model_no_inte)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##         Sum Sq Mean Sq NumDF   DenDF F value   Pr(>F)
## pasture 671.34 167.834     4  2.4376 57.1588 0.008596 **
## mineral   8.17   8.167     1 17.0000  2.7813 0.113684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
exactRLRT(mixed_model_no_inte)
```

```
##
##   simulated finite sample distribution of RLRT.
##
##   (p-value based on 10000 simulated values)
##
## data:
## RLRT = 22.247, p-value < 2.2e-16
```

Interpretation : From the ANOVA table, it is evident that the pasture is still a highly significant covariate which can be used in studying the milk yield across different blocks, but, mineral supplements is not significant. Also from the lmer() summary, we can see that all four of the pasture treatments are also significant. Also from the random LR test, we reject the null hypothesis at 5% level of significance and conclude that the variance of the random block effect is greater than 0.

Part (d)

```
anova(mixed_model,mixed_model_no_inte)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data8
## Models:
## mixed_model_no_inte: milk ~ pasture + mineral - 1 + (1 | block)
## mixed_model: milk ~ pasture * mineral - 1 + (1 | block)
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## mixed_model_no_inte    7 113.07 121.32 -49.537   99.074
## mixed_model           10 116.46 128.24 -48.232   96.464 2.6097  3     0.4558
```

Interpretation : Primarily, the AIC and BIC values from the ANOVA table show that the model with no interactions considered has both the values to be lower compared to the other model, which is indicative of the better fit to the data. Secondly, I feel that as we have seen previously that the interactions are not significant, it makes no sense in making the model more complicated. Lastly, I think that including interactions, when not needed also leads to difficulty in interpreting the coefficients of the model.

Part (e)

```r
par(mfrow=c(1,1))
qqnorm(residuals(mixed_model_no_inte))
qqline(residuals(mixed_model_no_inte),col="red")
```
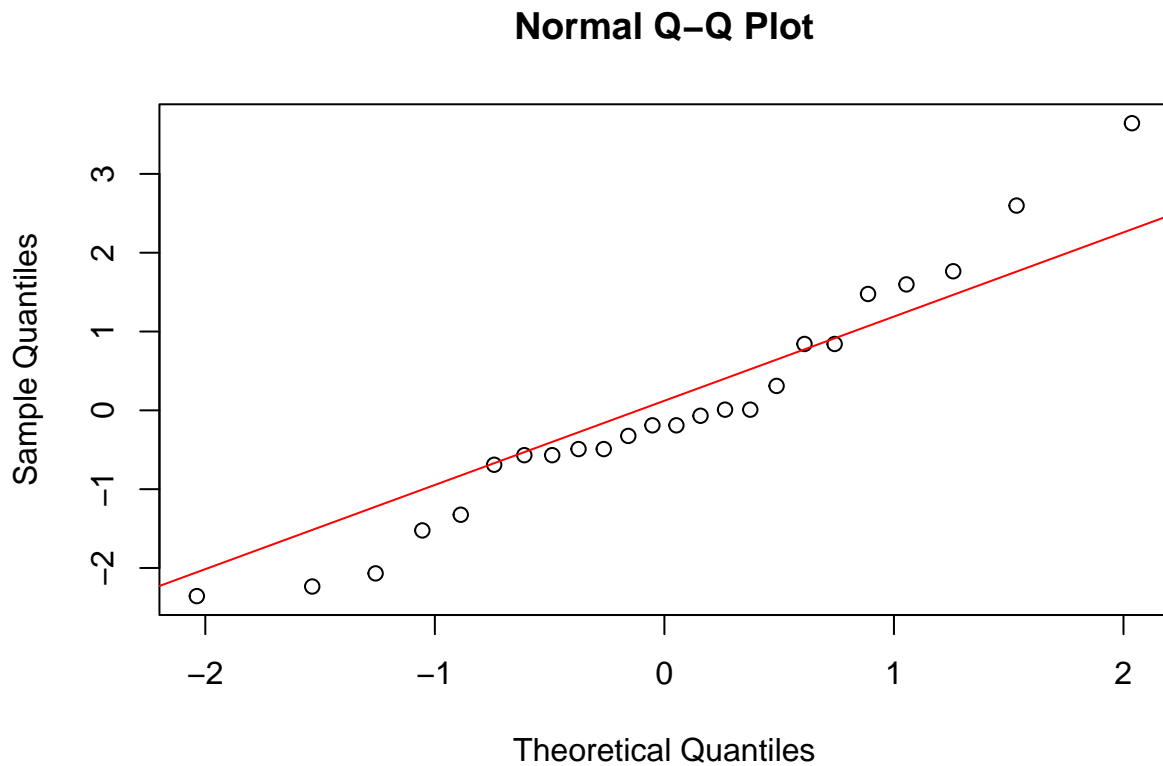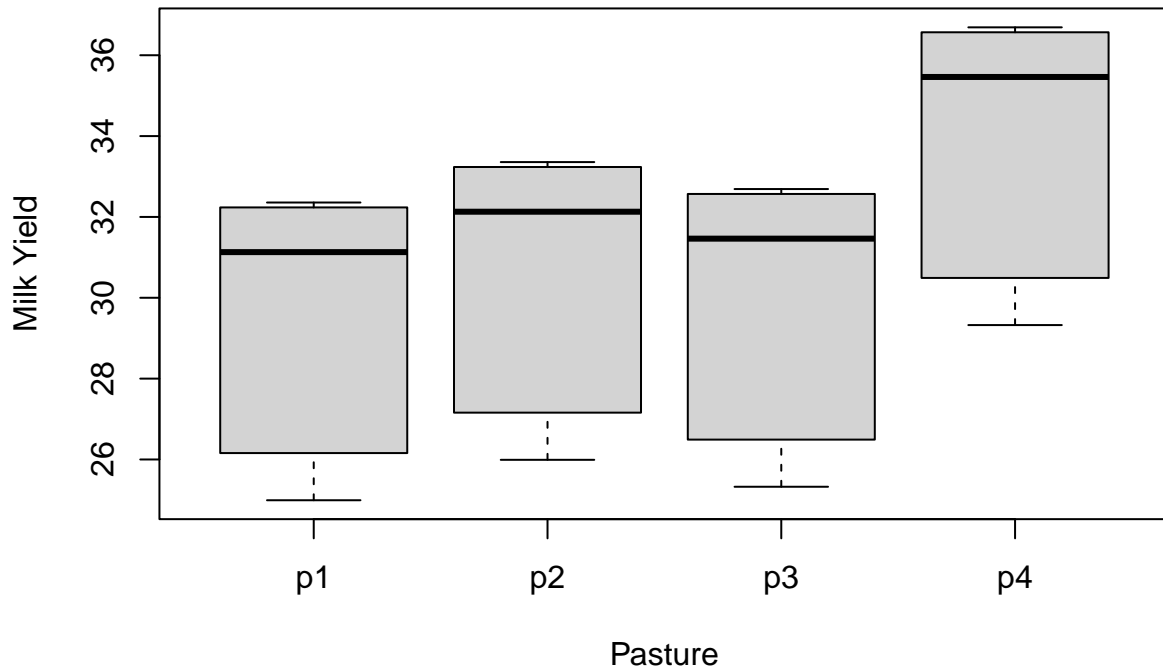
Normal Q–Q Plot


```r
plot(mixed_model_no_inte,xlab = "Fitted",ylab="Residuals",main="Residuals vs Fitted")
```

Residuals vs Fitted
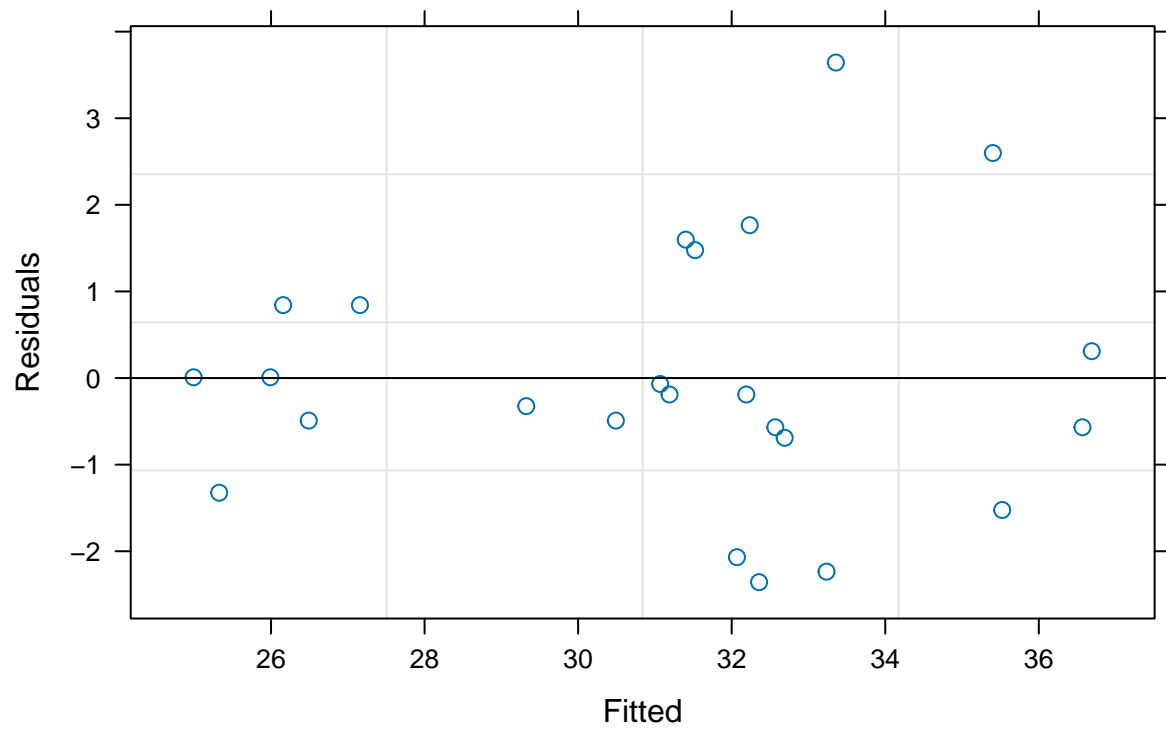

```
plot(data8$pasture,fitted(mixed_model_no_inte),xlab="Pasture",ylab="Milk Yield")
```

Conclusion : From the normal Q-Q plot, it can be observed that the residuals are distributed normally, which makes the assumption of normality hold.

From the residual vs fitted plot, it can be observed that the points do not follow any specific pattern, which signifies that the linearity assumption is valid. The points bounce from the 0 line forming a band, indicating the homogeneous error variances. Also, there does not seem to be any significant outlier.

From the histograms of milk yield on the basis of pasture levels, it can be evidently concluded that the median milk yield changes when moving from one pasture level to another, signifying the significance of all the four pasture treatments.

Part (f)

```
ea2(data8,design = 5,plot = 1)
```

Box plot for residuals
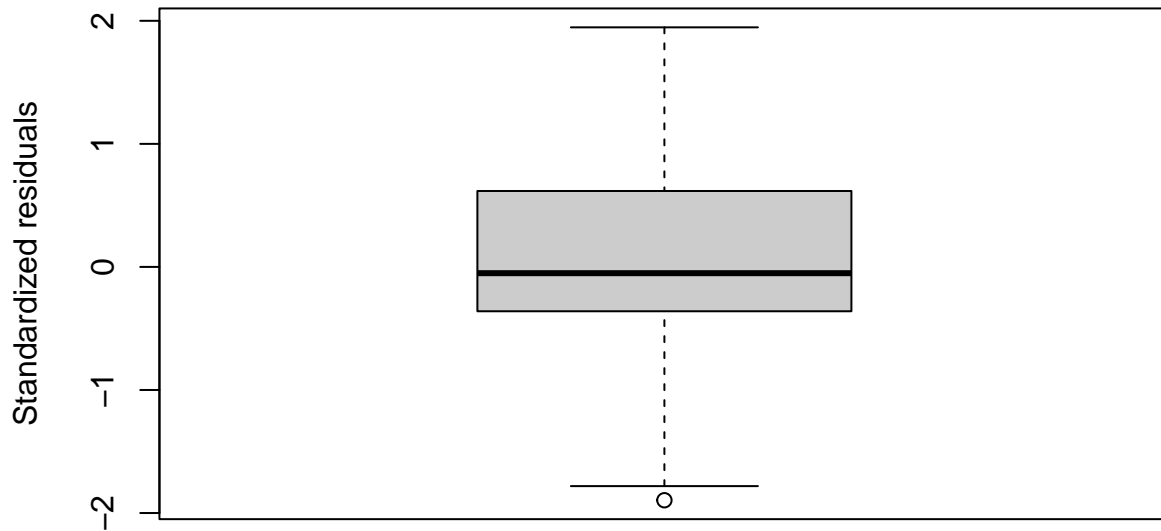

```
## $‘Marginal anova (Type III Sum of Squares)‘
##                numDF denDF   F-value p-value
## plot               3     6  5.456869  0.0377
## split.plot         1     8  3.629630  0.0932
## block              2     6 24.450479  0.0013
## plot:split.plot    3     8  0.864198  0.4981
##
## $‘Adjusted means (plot)‘
##   plot adjusted.mean standard.error tukey snk duncan t
## 1   p4       34.0000         0.8512     a   a      a a
## 2   p2       30.6667         0.8512    ab   b      b b
## 3   p3       30.0000         0.8512    ab   b      b b
## 4   p1       29.6667         0.8512     b   b      b b
##
## $‘Multiple comparison test (plot)‘
##      pair contrast p(tukey) p(snk) p(duncan)   p(t)
## 1 p4 - p2   3.3333   0.1135 0.0325    0.0325 0.0325
## 2 p4 - p3   4.0000   0.0587 0.0367    0.0185 0.0159
## 3 p4 - p1   4.3333   0.0427 0.0427    0.0144 0.0114
## 4 p2 - p3   0.6667   0.9421 0.5997    0.5997 0.5997
## 5 p2 - p1   1.0000   0.8385 0.6993    0.4517 0.4379
## 6 p3 - p1   0.3333   0.9918 0.7912    0.7912 0.7912
##
## $‘Adjusted means (split.plot)‘
##   split.plot adjusted.mean standard.error tukey snk duncan t
## 1         m2       31.6667         0.5243     a   a      a a
```

```
## 2           m1       30.5000          0.5243      a    a      a a
##
## $`Multiple comparison test (split.plot)`
##      pair contrast p(tukey) p(snk) p(duncan)    p(t)
## 1 m2 - m1   1.1667   0.0932 0.0932    0.0932 0.0932
##
## $`Adjusted means (plot in levels of split.plot)`
## $`Adjusted means (plot in levels of split.plot)`$`plot in  m1`
##   plot.split.plot adjusted.mean standard.error tukey snk duncan t
## 4           p4.m1       33.6667         1.0486     a   a      a a
## 3           p3.m1       30.0000         1.0486    ab   b      b b
## 2           p2.m1       29.3333         1.0486    ab   b      b b
## 1           p1.m1       29.0000         1.0486     b   b      b b
##
## $`Adjusted means (plot in levels of split.plot)`$`plot in  m2`
##   plot.split.plot adjusted.mean standard.error tukey snk duncan  t
## 8           p4.m2       34.3333         1.0486     a   a      a  a
## 6           p2.m2       32.0000         1.0486     a  ab     ab ab
## 5           p1.m2       30.3333         1.0486     a   b      b  b
## 7           p3.m2       30.0000         1.0486     a  ab      b  b
##
##
## $`Multiple comparison test (plot in levels of split.plot)`
## $`Multiple comparison test (plot in levels of split.plot)`$`plot in  m1`
##            pair contrast p(tukey) p(snk) p(duncan)    p(t)
## 1 p4.m1 - p3.m1   3.6667   0.1176 0.0301    0.0301 0.0301
## 2 p4.m1 - p2.m1   4.3334   0.0559 0.0330    0.0166 0.0133
## 3 p4.m1 - p1.m1   4.6667   0.0381 0.0381    0.0129 0.0088
## 4 p3.m1 - p2.m1   0.6667   0.9684 0.6614    0.6614 0.6614
## 5 p3.m1 - p1.m1   1.0000   0.9047 0.7826    0.5337 0.5134
## 6 p2.m1 - p1.m1   0.3333   0.9958 0.8261    0.8261 0.8261
##
## $`Multiple comparison test (plot in levels of split.plot)`$`plot in  m2`
##            pair contrast p(tukey) p(snk) p(duncan)    p(t)
## 1 p4.m2 - p2.m2   2.3333   0.4295 0.1427    0.1427 0.1427
## 2 p4.m2 - p1.m2   4.0000   0.0815 0.0488    0.0247 0.0200
## 3 p4.m2 - p3.m2   4.3333   0.0559 0.0559    0.0190 0.0133
## 4 p2.m2 - p1.m2   1.6667   0.6828 0.2840    0.2840 0.2840
## 5 p2.m2 - p3.m2   2.0000   0.5527 0.3980    0.2241 0.2034
## 6 p1.m2 - p3.m2   0.3333   0.9958 0.8261    0.8261 0.8261
##
##
## $`Adjusted means (split.plot in levels of plot)`
## $`Adjusted means (split.plot in levels of plot)`$`split.plot in  p1`
##   plot.split.plot adjusted.mean standard.error tukey snk duncan t
## 5           p1.m2       30.3333         1.0486     a   a      a a
## 1           p1.m1       29.0000         1.0486     a   a      a a
##
## $`Adjusted means (split.plot in levels of plot)`$`split.plot in  p2`
##   plot.split.plot adjusted.mean standard.error tukey snk duncan t
## 6           p2.m2       32.0000         1.0486     a   a      a a
## 2           p2.m1       29.3333         1.0486     a   a      a a
##
## $`Adjusted means (split.plot in levels of plot)`$`split.plot in  p3`
```

```
##   plot.split.plot adjusted.mean standard.error tukey snk duncan t
## 3          p3.m1            30         1.0486     a   a      a a
## 7          p3.m2            30         1.0486     a   a      a a
##
## $'Adjusted means (split.plot in levels of plot)'$'split.plot in  p4'
##   plot.split.plot adjusted.mean standard.error tukey snk duncan t
## 8          p4.m2       34.3333         1.0486     a   a      a a
## 4          p4.m1       33.6667         1.0486     a   a      a a
##
##
## $'Multiple comparison test (split.plot in levels of plot)'
## $'Multiple comparison test (split.plot in levels of plot)'$'split.plot in  p1'
##         pair contrast p(tukey) p(snk) p(duncan)  p(t)
## 1 p1.m2 - p1.m1   1.3333    0.308  0.308     0.308 0.308
##
## $'Multiple comparison test (split.plot in levels of plot)'$'split.plot in  p2'
##         pair contrast p(tukey) p(snk) p(duncan)   p(t)
## 1 p2.m2 - p2.m1   2.6667   0.0611 0.0611    0.0611 0.0611
##
## $'Multiple comparison test (split.plot in levels of plot)'$'split.plot in  p3'
##         pair contrast p(tukey) p(snk) p(duncan) p(t)
## 1 p3.m1 - p3.m2        0        1      1         1    1
##
## $'Multiple comparison test (split.plot in levels of plot)'$'split.plot in  p4'
##         pair contrast p(tukey) p(snk) p(duncan)   p(t)
## 1 p4.m2 - p4.m1   0.6666   0.6011 0.6011    0.6011 0.6011
##
##
## $'Residual analysis'
## $'Residual analysis'$values
##                                   values
## p.value Shapiro-Wilk test         0.5153
## p.value Bartlett test (plot)       0.8121
## p.value Bartlett test (split.plot)  0.4609
## p.value Bartlett test (plot*split.plot)  0.4044
## AIC                              106.1717
## BIC                              114.4794
## first value most discrepant       10.0000
## second value most discrepant      20.0000
## third value most discrepant       11.0000
## Mean Square of Error a             4.3472
## Mean Square of Error b             2.2500
## Coefficient of Variation a         6.7078
## Coefficient of Variation b         4.8257
##
## $'Residual analysis'$residuals
##          1           2           3           4           5           6
##  0.01570820 -0.31762513  0.61368477 -0.05298189  0.61368477 -0.05298189
##          7           8           9          10          11          12
##  0.59025559 -1.40974441 -0.30404685  2.02928648 -1.85676252  0.47657082
##         13          14          15          16          17          18
## -1.51171459  0.82161874  0.67252396 -0.32747604  1.24307774 -0.42358892
##         19          20          21          22          23          24
## -0.30963792 -1.97630459 -0.83732694  1.82933973  0.73722045 -0.26277955
```

```
## attr(,"label")
## [1] "Residuals"
##
## $`Residual analysis`$`standardized residuals`
##            1            2            3            4            5            6
##   0.01507063 -0.30473322  0.58877626 -0.05083144  0.58877626 -0.05083144
##            7            8            9           10           11           12
##   0.56629803 -1.35252507 -0.29170606  1.94692089 -1.78139941  0.45722755
##           13           14           15           16           17           18
## -1.45035644  0.78827051  0.64522726 -0.31418429  1.19262315 -0.40639610
##           19           20           21           22           23           24
## -0.29707020 -1.89608945 -0.80334114  1.75508966  0.70729781 -0.25211374
```

I find that using the *easyanova* package to perform the split-plot analysis is more informative compared to the previous approaches. The output consists of Analysis of Variance comparing the the the different types of design, v.i.z, plot, split-plot, blocking, etc. and compare their significance on the basis of the p-value. Also, we get adjusted means of the response variable on the basis of different kinds of pastures, minerals and their interactions, which would help to signify which treatment or treatment combinations are significant. It also provides with different kinds of testing of hypothesis for checking the assumptions of the residuals along with the residuals and standardized residual values, which in turn helps vastly in residual analysis. The results also include different plots for the residuals, which helps to verify the assumptions regarding the residuals graphically.

PROBLEM 4.

```r
#Loading the data


data_4 = read.table("C:/Users/Jayaditya Nath/Downloads/alzheim.dat", header=FALSE)
colnames(data_4) <- c("group", "t1", "t2", "t3", "t4", "t5")
data_4$group <- factor(data_4$group)
data_4$subject <- 1:nrow(data_4)
data_4 <- data_4[,c(7,1:6)]
data_4$subject <- factor(data_4$subject)


data_4 <- data.frame(
  score = as.vector(unlist(data_4[,3:7])),
  visit = c(rep(1, nrow(data_4)), rep(2, nrow(data_4)),
            rep(3, nrow(data_4)), rep(4, nrow(data_4)), rep(5, nrow(data_4))),
  group = rep(data_4$group, times = 5),
  subject = rep(data_4$subject, times = 5))
```

Part (a)

```r
# Plot profiles for each subject
plot_obs = xyplot(score ~ visit | group,groups = group , data_4, as.table = TRUE, type = c("p","l", "sm
plot_obs
```
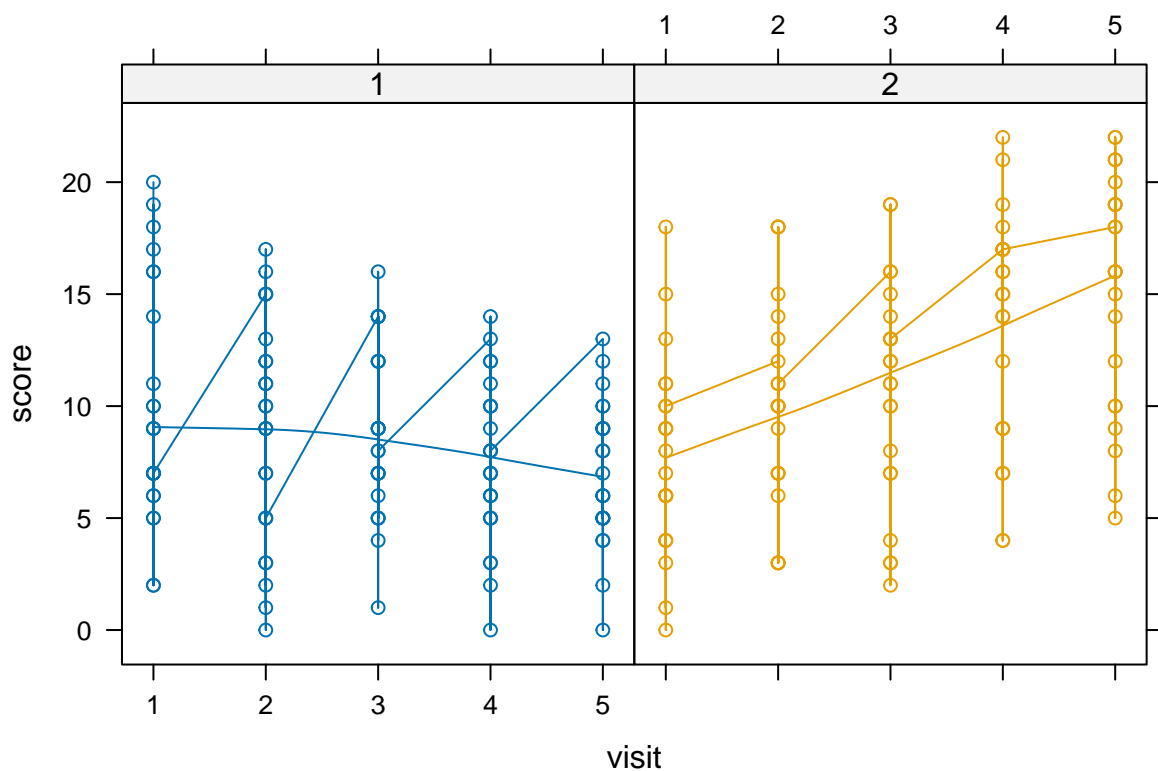
```r
# Plotting mean score for each group versus visit
mean_scores <- aggregate(score ~ visit + group, data = data_4, FUN = mean)
ggplot(data = mean_scores, aes(x = visit, y = score, group = group, color = group)) +
  geom_line() +
  geom_point() +
  labs(x = "Visit", y = "Mean Score", color = "Group") +
  ggtitle("Mean Score for Alzheimer's Patients by Group") +
  theme_minimal()
```

Mean Score for Alzheimer's Patients by Group
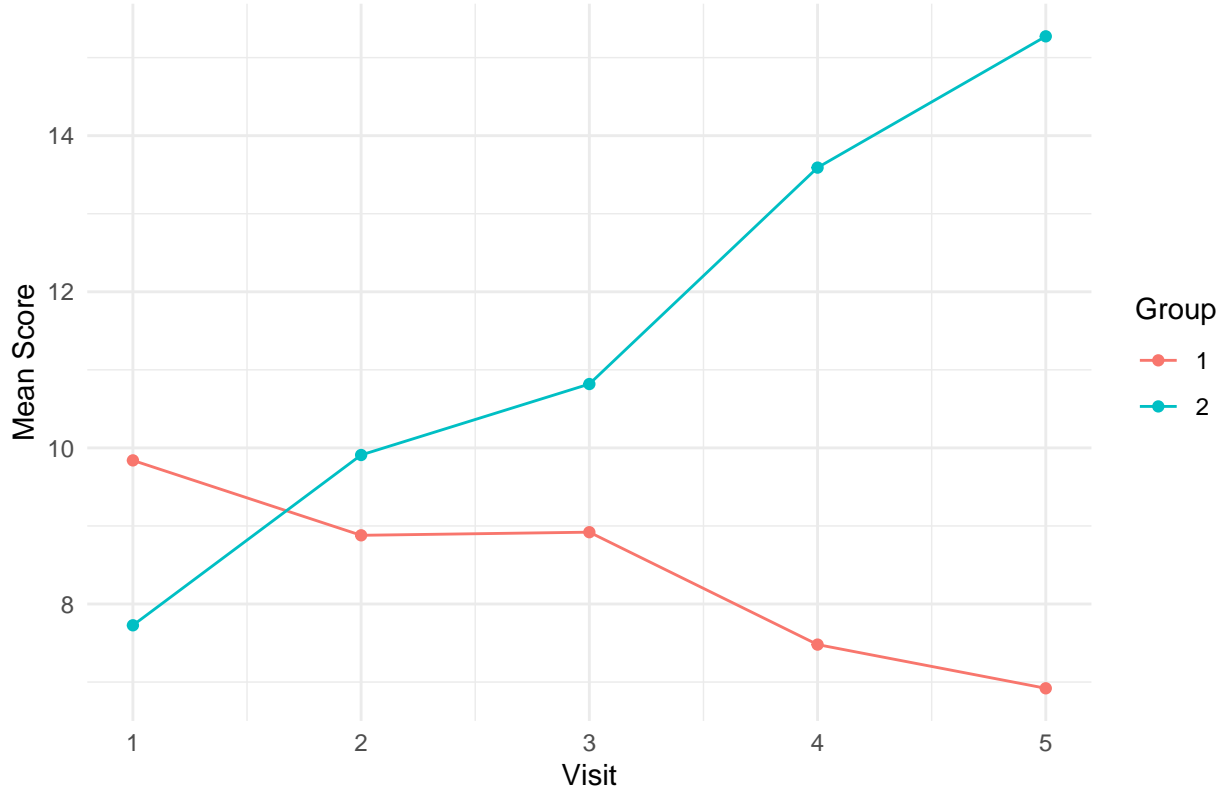
Group
- 1
- 2

Interpretation : From the plot, we can observe that the trend of the cognitive test score sees a decay with the number of visits for the 1st group receiving placebo, but, the opposite happens for the second group receiving lecithin as the trend in the cognitive score seems to experience a sharp rise, indicative of the fact that the patients are being able to remember more words compared to the placebo group. This in turn shows that the lecithin drug was indeed effective in slowing or even halting the memory impairment in Alzheimer's patients.

Part (b)

The model used for the given longitudinal study is :

$Y_{ijk} = \beta_0 + \beta_1 g_i + \beta_2 v_j + b_{0k} + \epsilon_{ijk}$, i = 1(|)2, j = 1(|)5 & k = 1(|)47, $\epsilon_{ijk} \sim$ i.i.d $N(0, \sigma^2)$ and $b_{0k} \sim N(0, \sigma_b^2)$

where, Y_{ijk} is the response corresponding to the ith group, jth visit and kth subject

β_0 is the fixed intercept.

β_1 is the fixed effect coefficient due to the ith group.

β_2 is the fixed effect coefficient due to the jth visit.

g_i denotes the ith group.

v_j denotes the jth visit.

b_{0k} is the random intercept due to the kth subject

ϵ_{ijk} is the random error.

Part (c)

The implied covariance structure of repeated visits on the kth subject of the ith group is :

$$\Sigma_k = \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma^2 & . & . & . & \sigma^2 \\ \sigma^2 & \sigma_b^2 + \sigma^2 & . & . & . & \sigma^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma^2 & \sigma^2 & . & . & . & \sigma_b^2 + \sigma^2 \end{pmatrix}$$

Part (d)

```
model_4 = lmer(score ~ visit + group - 1 + (1|subject),data = data_4,REML = F)

summary(model_4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: score ~ visit + group - 1 + (1 | subject)
##    Data: data_4
##
##      AIC      BIC   logLik deviance df.resid
##   1281.7   1299.0   -635.9   1271.7      230
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7377 -0.6083 -0.1167  0.5834  3.1157
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 15.128   3.890
##  Residual              8.246   2.872
## Number of obs: 235, groups:  subject, 47
##
## Fixed effects:
##         Estimate Std. Error       df t value Pr(>|t|)
## visit     0.4936     0.1325 188.0000   3.727 0.000257 ***
## group1    6.9271     0.9105  70.7403   7.608 9.08e-11 ***
## group2    9.9828     0.9594  67.7521  10.405 1.10e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        visit  group1
## group1 -0.436
## group2 -0.414  0.181
```

Here, we fit the linear mixed model using the Maximum Likelihood approach. We observe that the fixed effect coefficients are significant. Also, $\hat{\sigma}^2 = 8.246$ and $\hat{\sigma_b}^2 = 15.128$.

Part (e)

The model used for the given longitudinal study is :

$Y_{ijk} = \beta_0 + \beta_1 g_i + (\beta_2 + b_{1k})v_j + b_{0k} + \epsilon_{ijk}$, i = 1(|)2, j = 1(|)5 & k = 1(|)47, $\epsilon_{ijk} \sim$ i.i.d $N(0,\sigma^2)$ and $b_{0k} \sim N(0,\sigma_b^2)$

where, Y_{ijk} is the response corresponding to the ith group, jth visit and kth subject

β_0 is the fixed intercept.

15

β_1 is the fixed effect coefficient due to the ith group.

β_2 is the fixed effect coefficient due to the jth visit.

g_i denotes the ith group.

v_j denotes the jth visit.

b_{0k} is the random intercept due to the kth subject

b_{1k} is the random coefficient due to the jth visit

ϵ_{ijk} is the random error.

Part (f)

```
model_4_e = lmer(score ~ group - 1 + (1 + visit|subject),data = data_4,REML = F)
summary(model_4_e)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: score ~ group - 1 + (1 + visit | subject)
##    Data: data_4
##
##      AIC      BIC   logLik deviance df.resid
##   1200.0   1220.7   -594.0   1188.0      229
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.65541 -0.43601  0.05944  0.43767  2.28266
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subject  (Intercept) 41.339   6.430
##           visit        2.301   1.517   -0.78
##  Residual              3.104   1.762
## Number of obs: 235, groups:  subject, 47
##
## Fixed effects:
##         Estimate Std. Error       df t value Pr(>|t|)
## group1    8.2080     0.8179  47.0000   10.04 2.86e-13 ***
## group2   11.9823     0.8719  47.0000   13.74  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        group1
## group2 0.000
```

Interpretation : For the given model, the fixed effect(i.e, group) seems to be statistically significant. So, we can finally conclude that the cognitive test score for the subjects receiving Lecithin drug should have a higher cognitive test score compared to the ones in the Placebo group.

Part (g)

```
anova(model_4,model_4_e)
```

```
## Data: data_4
## Models:
## model_4: score ~ visit + group - 1 + (1 | subject)
## model_4_e: score ~ group - 1 + (1 + visit | subject)
##         npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## model_4     5 1281.7 1299.0 -635.86   1271.7
## model_4_e   6 1200.0 1220.7 -593.98   1188.0 83.76  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation : As the p-value is less than the significance level of $\alpha = 0.05$, we reject the null hypothesis at 5% level of significance and can thus conclude that the model with the random intercept and random coefficient for visit is a better fit compared to the model with the random intercept only. The AIC values from the ANOVA table also comply with our conclusion.
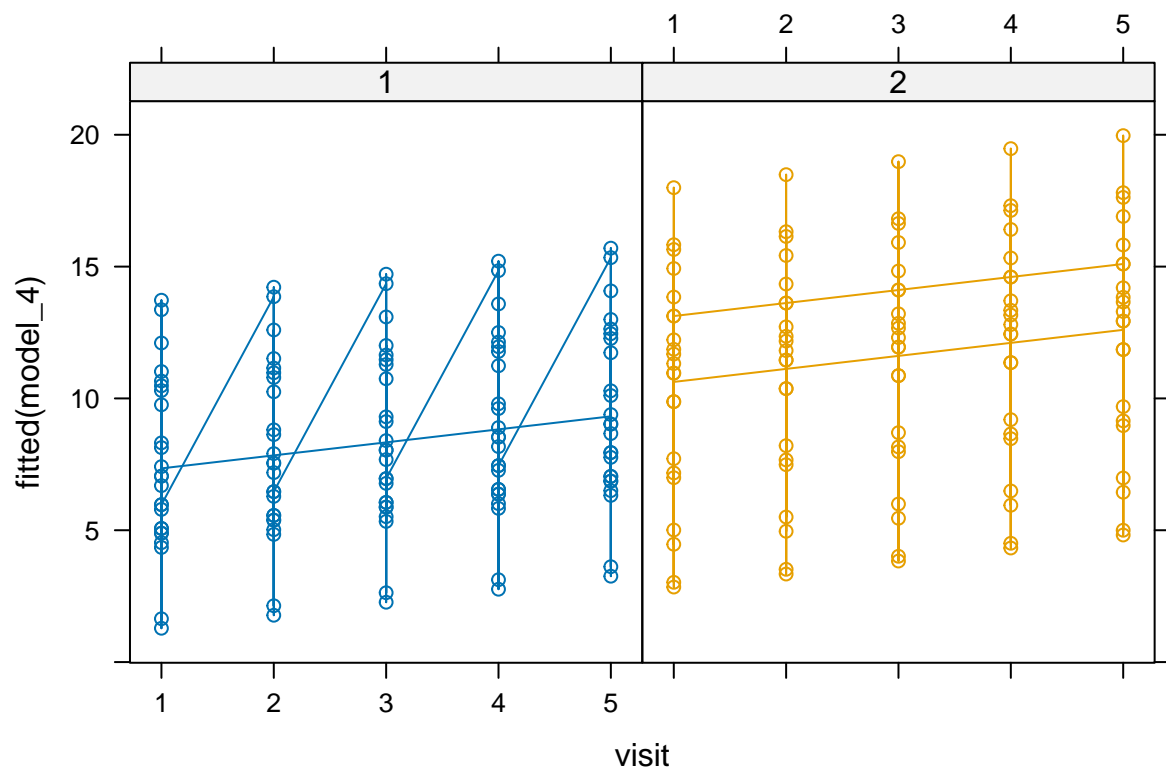
Part (h)

```
# Plot profiles for each subject
plot_rand_int = xyplot(fitted(model_4) ~ visit | group,groups = group , data_4, as.table = TRUE, type =

# Plot profiles for each subject
plot_mod_final = xyplot(fitted(model_4_e) ~ visit | group,groups = group , data_4, as.table = TRUE, typ

plot_obs
```
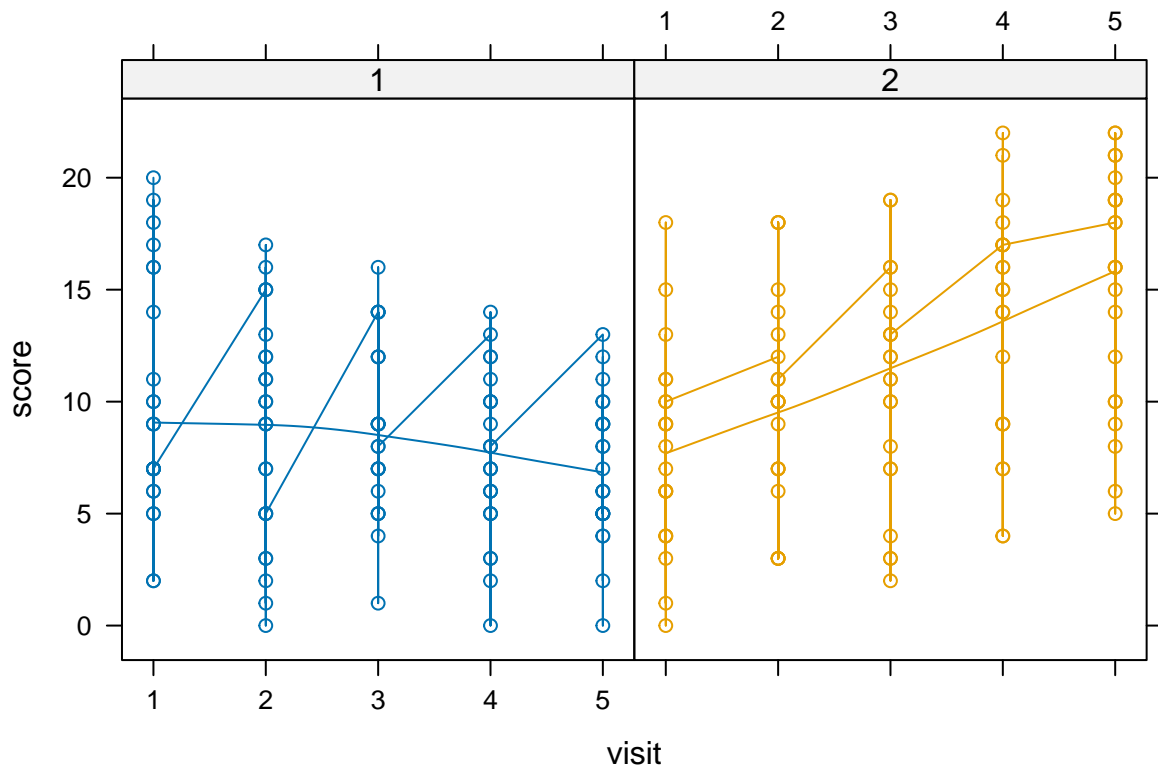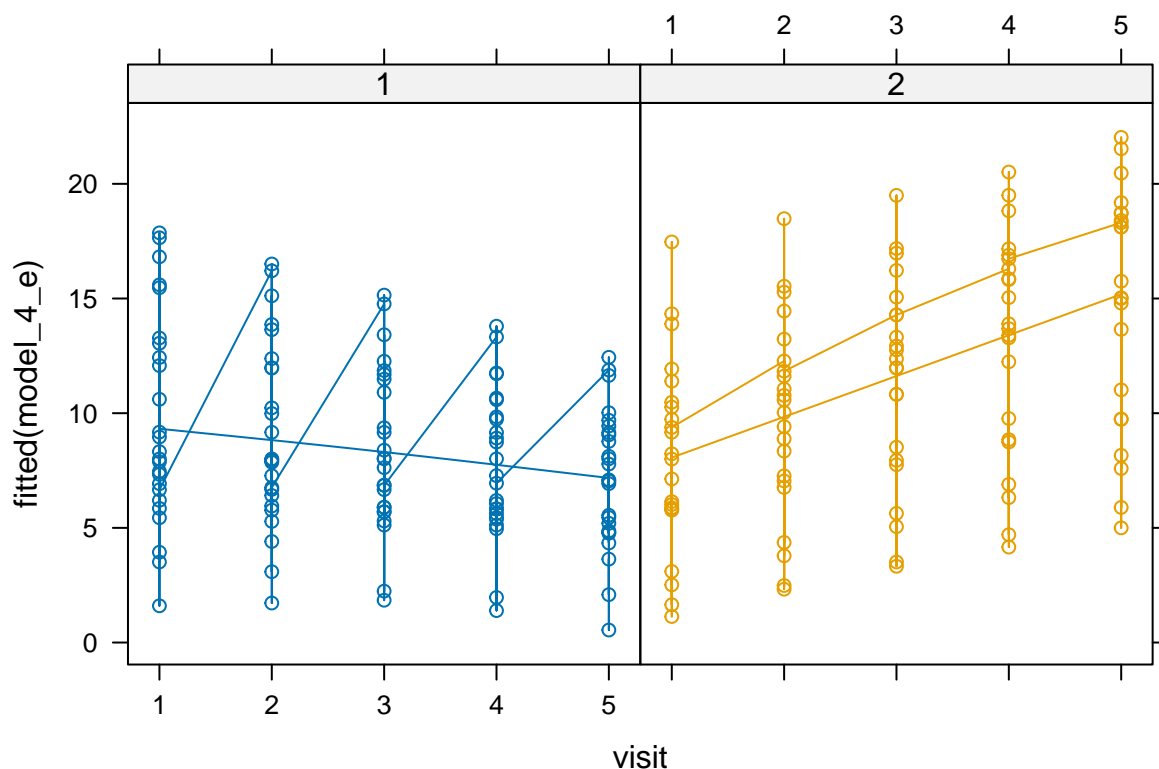
Interpretation : From the comparison of the graphs, we can observe that the random intercept model displays similar profiles for the subjects of both the groups, which we know is not the case from the observed data. On the other hand, adding a random effect to the *visit* variable shows that the model fits the data quite well as there is an increasing trend in the cognitive score for the group receiving Lecithin compared to the one receiving Placebo.

Part (i)

```
model_4_AR1 <- lme(score ~ visit + group, random = ~1|subject, correlation = corAR1(form = ~1|subject),
summary(model_4_AR1)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data_4
##        AIC      BIC    logLik
##    1214.719 1235.477 -601.3596
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev: 0.0007234615 4.996681
##
## Correlation Structure: AR(1)
##  Formula: ~1 | subject
##  Parameter estimate(s):
##       Phi
## 0.8307846
```

```
## Fixed effects:   score ~ visit + group
##                Value Std.Error  DF  t-value p-value
## (Intercept) 6.905598 1.0273470 187 6.721777  0.0000
## visit       0.494617 0.1876468 187 2.635891  0.0091
## group2      3.098284 1.2561012  45 2.466588  0.0175
##  Correlation:
##        (Intr) visit
## visit  -0.548
## group2 -0.572  0.000
##
## Standardized Within-Group Residuals:
##         Min         Q1         Med         Q3         Max
## -2.10109439 -0.69693214 -0.08009599  0.70507527  2.52163116
##
## Number of Observations: 235
## Number of Groups: 47
```

```r
model_4_unstruc <- lme(score ~ visit + group, random = ~1|subject, correlation = corSymm(form = ~1|subj
summary(model_4_unstruc)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data_4
##        AIC      BIC    logLik
##    1179.167 1231.061 -574.5836
##
## Random effects:
##  Formula: ~1 | subject
##         (Intercept) Residual
## StdDev:    4.209627 3.123693
##
## Correlation Structure: General
##  Formula: ~1 | subject
##  Parameter estimate(s):
##  Correlation:
##   1      2      3      4
## 2  0.259
## 3  0.300  0.699
## 4 -0.329  0.406  0.570
## 5 -0.727  0.114  0.323  0.843
## Fixed effects:   score ~ visit + group
##                Value Std.Error  DF  t-value p-value
## (Intercept) 6.750256 1.0194349 187 6.621566  0.0000
## visit       0.483863 0.1836953 187 2.634052  0.0091
## group2      3.525241 1.2491564  45 2.822097  0.0071
##  Correlation:
##        (Intr) visit
## visit  -0.545
## group2 -0.574  0.000
##
## Standardized Within-Group Residuals:
##         Min         Q1         Med         Q3         Max
## -2.50221070 -0.71701456 -0.02203092  0.62069979  3.61576746
##
## Number of Observations: 235
```

```
## Number of Groups: 47
```

Interpretation : The lower AIC value for the unstructured covariance model has a better predictive ability due to the lower AIC value compared to the model having AR(1) covariance structure.