UNIVERSITÉ DE GENÈVE

Université de Genève

S411014: Mixed Linear Models

Chemotherapy Analysis

GROUP

(20-345-278) Chen, Jiahao
(15-331-523) Dakoli, Isei
(16-465-321) Dantoni, Dario
(20-336-988) González, Marc
(15-347-826) Kalemi, Klea
(19-433-754) Motie, Shiva
(20-345-146) Roudaki, Pouya
(16-313-231) Sallin, Gabriel Jérôme
(11-331-527) Souvairan, Kia Judith

PROFESSOR

Dominique-Laurent Couturier

Research Center for Statistics
GSEM

Geneva, 14.01.2021

1 Variable Description

The dataset comprises five variables; `tumour`, `month`, `line`, `sensitivity` as observed variables, and `patient` as an unobserved variable. The response variable is `tumour` and is of class numeric. The variable `line` comprises 5 levels which will be treated as such for question 1, and considered a 3 levels factor for question 2. This makes sense in the estimation since levels 4 and 5 contain only 10% and 5% of the observations respectively. Variables `months` and `sensitivity` are regressors. Variable `patient` is a random effect factor because, while it provides the measurements, it is not of direct interest. `patient` is an independent experiment unit. Within `patient`, there are different number of treatment lines, and each line can last 3 months or less. This setup makes the dataset unbalanced and dictates the form that the variable `line` takes, a fixed effect, given that it is also of direct interest for the study's purpose. `line` and `sensitivity` are nested within `patient`, while `month` is nested within `line`. The summary of the dataset is presented in Listings (1) and (2).

2 Exploratory Data Analysis

Plots of the distribution of the response variable `tumour` against `line` (Figure 1) show that the the data is skewed to the left, more so for levels 2 and 3 of `line`. Given that the distribution of the response is not symmetric, we apply different transformations to the data (Figure 2) so as to respect the assumption of normality of the response. However, the skewness persists (Figure 2b and 2c) and so do the fat tails (Figure 2d). We keep the `tumour` untransformed in the estimation.

It appears that there are equal variances between `tumour` and `line` with 3 levels but not for `line` with 5 levels. See Figure 4. Heteroscedasticity is detected for both `month` and `sensitivity`, as in Figure 5. These plots also show that there is a potential linear relationship between the response and the covariate `sensitivity`, but a quadratic relationship with `month`.

Figures 6 and 7 indicate that we should consider interactions between `line` and `month`, as well as `line` and `sensitivity`. Indeed we can observe how the slope for `line 1` is different than the slope for the subsequent treatment lines. Based on Figure 3a, we would consider a random intercept per `patient`. Meanwhile, Figure 8 indicates different slopes of `sensitivity` per `patient`, making of it a potential random slope. In Figure 9 the interaction plot indicates that we should consider modelling the interaction between `patient` and `line`. This is as well supported by the fact that `tumour` per `patient` is different per level of `line`. We also consider the interaction between `patient` and the regressor `month` as in Figure 10. Some potential outliers can be detected in Figures 4a and 5a.

3 Model Building and Hypothesis

The research question is about whether `sensitivity` can be used to predict `tumour`, and whether `sensitivity` decreases with time and is stronger for line 1 than for lines 2 and 3+. A first model, containing all possible interactions established from exploratory analysis:

$$\begin{aligned} T_{i(j(m),s)} = \quad & \beta_0 + \beta_1 \cdot s + \beta_2 \cdot m + (\beta_3 + \beta_4 \cdot s + \beta_5 \cdot m) \cdot L_2 + (\beta_6 + \beta_7 \cdot s + \beta_8 \cdot m) \cdot L_3 \\ & + (\beta_9 + \beta_{10} \cdot s + \beta_{11} \cdot m) \cdot L_4 + (\beta_{12} + \beta_{13} \cdot s + \beta_{14} \cdot m) \cdot L_5 \\ & + (\gamma_{0,i} + \gamma_{1,i} \cdot s + \gamma_{2,i} \cdot m + (\gamma_3 \cdot L)_{i(j)}) + \epsilon_{i(j(m),s)} \end{aligned} \tag{1}$$

for the i^{th} patient, j^{th} line, s^{th} sensitivity and m^{th} month, with $\gamma_0 \sim \mathcal{N}(0, \sigma^2_{\gamma_0})$, $\gamma_1 \sim \mathcal{N}(0, \sigma^2_{\gamma_1})$, $\gamma_2 \sim \mathcal{N}(0, \sigma^2_{\gamma_2})$, $\gamma_3 \sim \mathcal{N}(0, \sigma^2_{\gamma_3})$, $\epsilon_{i(j(m),s)} \sim \mathcal{N}(0, \sigma^2_{\epsilon})$.

Listing 3 summarizes this first model. The normal QQ plot of the residuals indicate a fat left tail and thus the normality assumption does not seem to be satisfied. See Figure 11. The `patient` random intercept, the `sensitivity` random slope and the `month` random slope satisfy the normality assumption, while the other random effects do not. See Figure 12. All the plots in Figure 13 show heteroscedasticity. Furthermore, we detect linearity in "Residual VS Sensitivities" and "Residual VS Line", while "Residuals VS Month" presents a quadratic shape. The fixed effects parameter estimates suggest that `month` is not significant, which could relate to the non-linearity. Since the coefficients for the interactions between fixed effects are mostly not significant and given the lack of variability of `tumour` per `line` throughout time, see Figure 7, we doubt the presence of the interaction ($line * month$).

Regarding the random effects, the correlation between the random intercept and the random slope for `sensitivity` is -1 (as in Figure 14), a result of fitting a too complex model to a small dataset. Considering then independence between the random intercept and the random slope leads to a random slope variance of 0. In addition, the variance of the other random effects is close to 0, a hint to not consider them in further models. Then we test for the interaction between `month` and `line` in the fixed effect part. See Listings (4). The p-value is very close to 5% and the AIC is slightly lower for the model without the interaction so we decide to consider it as the better model. It is of interest for the second research question to preserve the interaction between `line` and `sensitivity`, so it will be kept. Beforehand, we will address the quadratic relation of `tumour` and `month`.

Moreover, we noticed through the model check above that heteroskedasticity is present. To take care of this we modelled the different variances for each level of the factor `line`. We test for this extension in Listing 6, and there is not enough evidence to reject the full model. Then we tried to model the variance as a function of the exponential value of `sensitivity` assuming thus that $\epsilon_{i(j(m),s)} \sim \mathcal{N}(0, \sigma^2_\epsilon exp(2\delta s_i))$. We test for this extension, $H_0 : \delta = 0$, and we have enough evidence to reject the full model (p-value of 0.0398 and a lower AIC). See Listing 7. Finally, we repeated the process explained above for `line` with 3 levels. In both cases the model considering the heteroscedasticity is better. See Listing 6 and 7. We also tried to boost our model by modelling correlated residuals with ARMA(2,1), between cluster correlation, and robustness using Huber weights. Appendix A.5: Additional remarks, elaborates on why these models were not considered.

Our final model is:

$$\begin{aligned} T_{i(j(m),s)} = \quad & \beta_0 + \beta_1 \cdot s + \beta_2 \cdot m + \beta_2 \cdot m^2 + (\beta_3 + \beta_4 \cdot s) \cdot L_2 + (\beta_6 + \beta_7 \cdot s) \cdot L_3 \\ & + (\beta_9 + \beta_{10} \cdot s) \cdot L_4 + (\beta_{12} + \beta_{13} \cdot s) \cdot L_5 + (\gamma_{0,i} + \epsilon_{i(j(m),s)}) \end{aligned} \tag{2}$$

where $\gamma_0 \sim \mathcal{N}(0, \sigma^2_{\gamma_0})$, $\epsilon_{i(j(m),s)} \sim \mathcal{N}(0, \sigma^2_\epsilon exp(2\delta s_i))$.

Given the 2 research questions, the first test is about the coefficient of `sensitivity`; thus the hypothesis tested is $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 > 0$ (and where the choice of contrast for fixed effect factors is not relevant to our question). For the second question, `line_3` with 3 levels (1, 2, and 3+) is used. To answer it, we use treatment contrast for `line_3` with `line 1` as reference. The hypotheses are $H_0 : \beta_4 = 0$ *and* $\beta_7 = 0$; $H_1 : \beta_4 < 0$ *and* $\beta_7 < 0$. Graphically, it corresponds to testing in Figure 6b if the slope of `line 1` is significantly higher than the slope of `line 2` and `line 3`.

4 Model Estimation, Check and Hypothesis Testing

We made a diagnostic analysis to determine the validity of the model with five levels of `line`. To check that the residuals are indeed normally distributed, we simulated 8 QQ-plots with data following a $\mathcal{N}(0, \sigma^2_\epsilon)$ and hid the QQ-plot of the model residuals in Figure 15a among the simulated (see Figure 16). As it is not evident where the one with our model residuals is hidden, we can conclude that the residuals seem to satisfy the normality assumption. In Figure 15b we note that it is reasonable to say that the random intercept follows a Normal distribution since we have only 19 clusters at the patient level. The interquartile range of the plots "Residuals VS Line" and "Residuals VS Month" in Figure 17 still show heteroscedasticity for `line`, while the assumption of linearity is satisfied (see both plots). Furthermore, some outliers are still present. The same comments apply to the model with 3 levels of `line`. See Figures 18, 19.

Although we consider a heteroscedastic model, since the dataset is unbalanced, we tested the signifance of the interactions $(sensitivity * line)$ with the LRT corrected by Bootstrap (See Figure 21) in the model without heteroscedasticity (this is because `lmer` was used for the simulations, and does not allow to model for heteroscedasticity). The coefficients are jointly significant (See Listing 5, the p-value is 0.03).

5 Conclusion

Q1: Can the sensitivity score be used to predict the treatment effect on the patient at the hospital?

Listing 8 shows the output of the final model (for `line` with 5 levels) with a confidence level of 5%. A high sensitivity score, indicated by the positive coefficient, implies that the tumour is resistant to the drug (tumour marker is not shrinking). In this case we have statistical evidence that `sensitivity` is a good predictor for `tumour`.

Q2: Does the sensitivity score decrease with time and is it stronger for line 1 than for lines 2 and 3+?

The variable `sensitivity` is a better predictor for `line 1` compared to `line 2` as the coefficient β_4 is negative and significant (p-value for one sided test is $(0.0752/2) < 0.05$). The variable sensitivity is a better predictor for `line 1` compared to `line 3` as the coefficient β_7 is negative and significant (p-value for one sided test is $(0.0027/2) < 0.05$). See Listing 10 for both.

Given that we are modelling heteroscedasticity and random effects, we are limited to the package `nlme` thus the only joint test that we can perform is through the anova function. The interaction coefficients in the model where `line` has 3 levels are jointly significant as the p-value is 0.0085. See Listing 9. We can conclude that the `sensitivity` score decreases with time (given the way the experiment is conducted: moving from line 1 to 3+ time increases) and works best for initial chemotherapy lines of treatment.

A Appendix

A.1 First look

```
> summary(df)[,1:3]
   patient        tumour          line_3
A       :14   Min.   :-3.8546   line 1  :54
B       :12   1st Qu.:-1.2685   line 2  :36
G       :12   Median :-0.5630   lines 3+:47
H       :11   Mean   :-0.7329
K       :10   3rd Qu.:-0.1106
F       : 9   Max.   : 1.1342
(Other):69
```

Listing 1: Variable summary 1

```
> summary(df)[,4:6]
     month         sensitivity         line
Min.   :1.000   Min.   :0.003248   line 1:54
1st Qu.:1.000   1st Qu.:0.377984   line 2:36
Median :2.000   Median :0.563378   line 3:26
Mean   :1.876   Mean   :0.565992   line 4:14
3rd Qu.:3.000   3rd Qu.:0.782861   line 5: 7
Max.   :3.000   Max.   :1.213053
```

Listing 2: Variable summary 2

Code outputs as in Listings (1) and (2) describe the shape of the data after the pertinent adaptations.

A.2 Exploratory Data Analysis

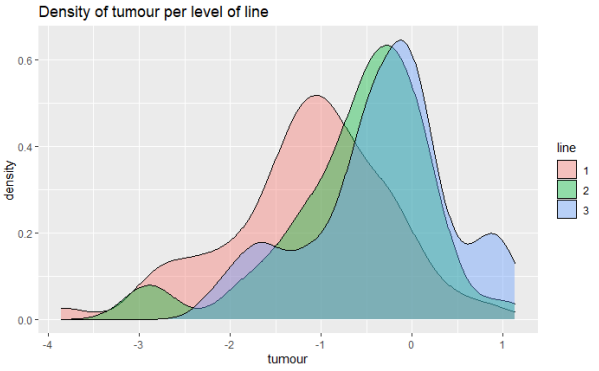
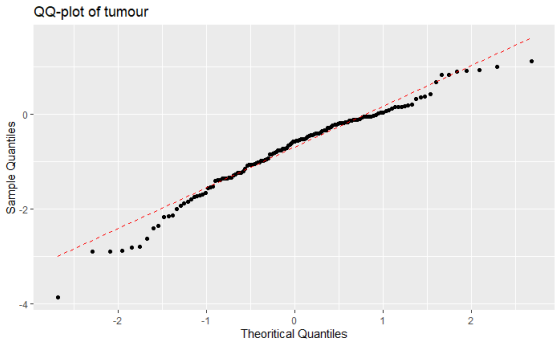
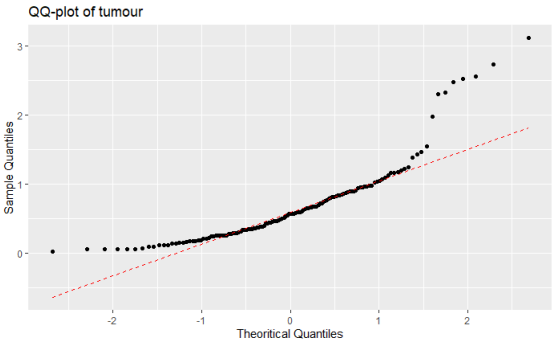


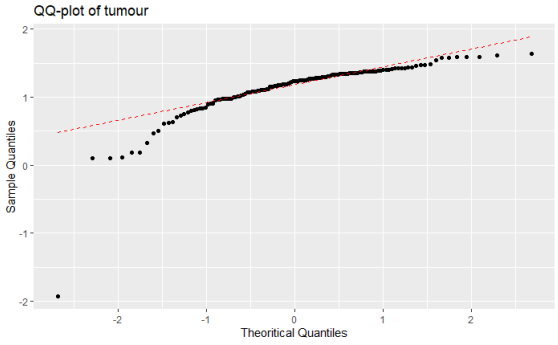
Figure 1: pdf of tumour size (for 3 levels of line)



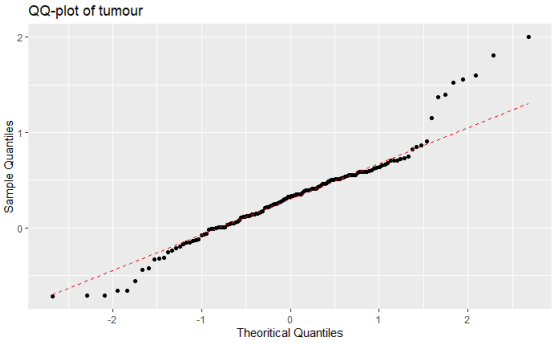
(a) Unmodified distribution



(b) Exponential transform



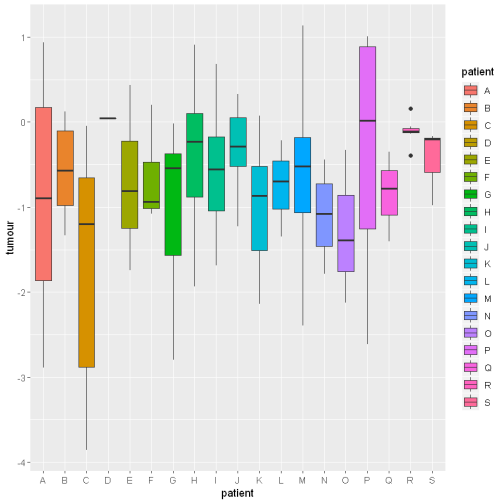
(c) Log transform



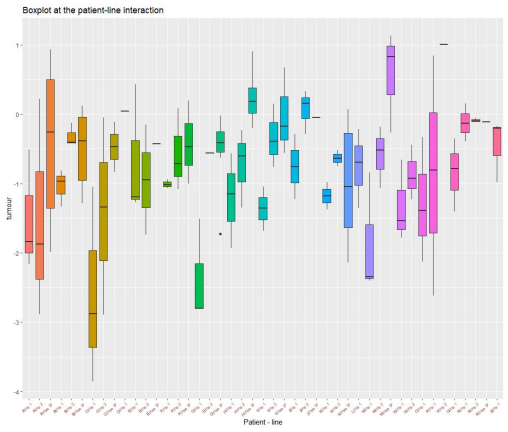
(d) Hyperbolic arc-tangent transform

Figure 2: QQ-plots of `tumour` for the original variable and some potential transformations

The distributional properties of the response variable, `tumour`, are displayed in Figure (1) by level of `line`. For lines 2 and 3+, the empirical density shows skewness to the left. Some transformations are considered and exposed in the QQ-plots (Figure (2)) using the quantiles of a normal: (a) shows the untransformed version, while figures (b) and (c) correspond to the classical transformations employed to regulate the heteroscedasticity, although in this case they fail to adapt the variable to a normal distribution. Plot (d) suggests a more refined case with a well behaved centre but big fat tails.



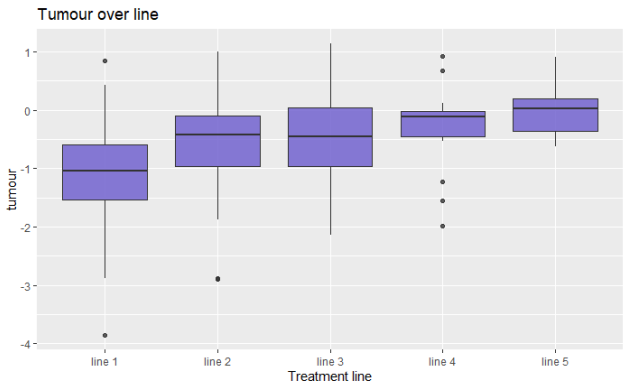
(a) Boxplot per patient



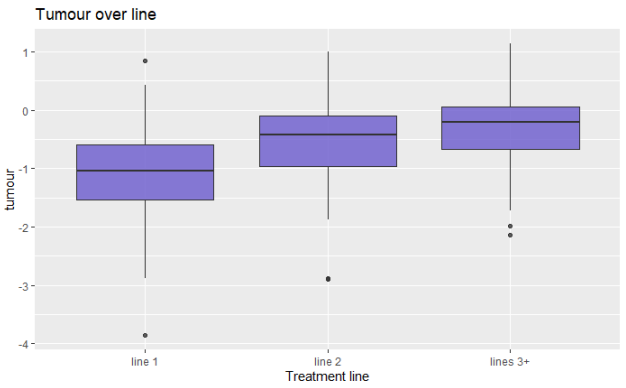
(b) Boxplot per level of `line` within `patient`

Figure 3: Boxplots of the `patient` variable

In Figure 3 some skewness is observed both in (a).

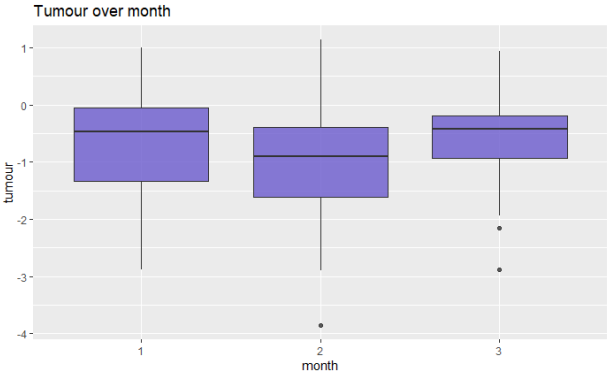


(a) Tumour per line (5 levels)

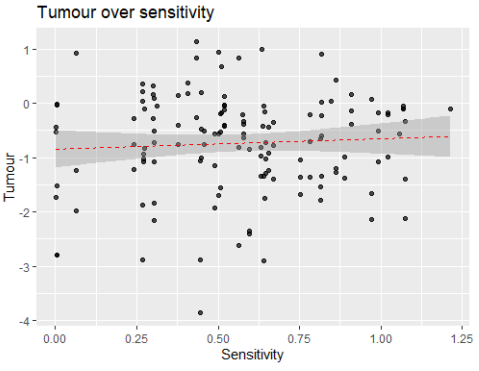


(b) Tumour per line (3 levels)

Figure 4: Boxplots of tumour per line

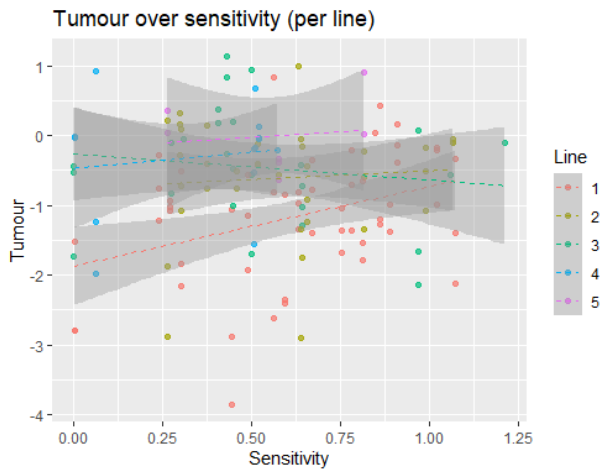


(a) Boxplot of tumour per month

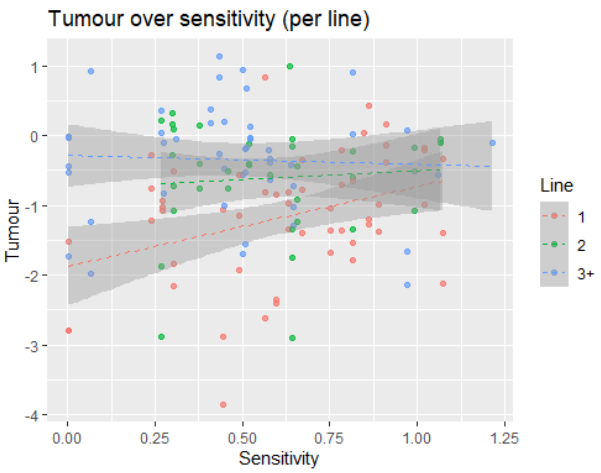


(b) Plot of tumour and sensitivity

Figure 5: Linearity and heteroscedasticity check

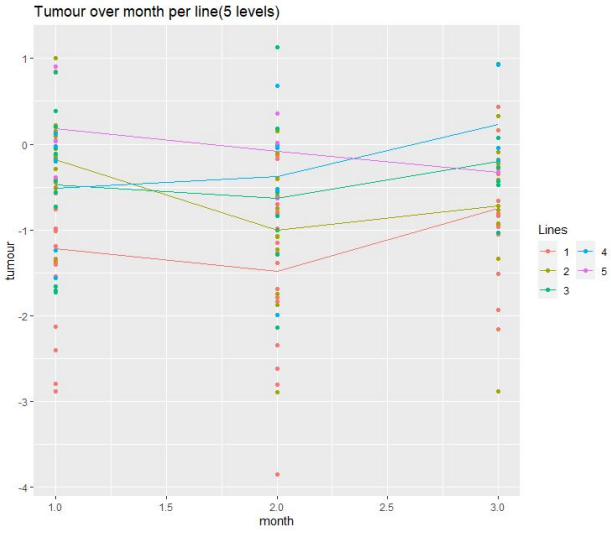


(a) Tumour VS Sensitivity per line (5 levels)

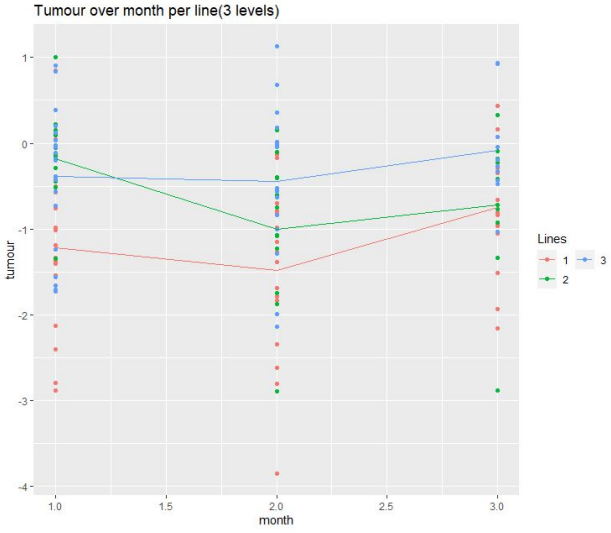


(b) Tumour VS Sensitivity per line (3 levels)

Figure 6: Interactions between sensitivity and line



(a) Tumour VS Month per line (5 levels)



(b) Tumour VS Month per line (3 levels)

Figure 7: Interactions between month and line

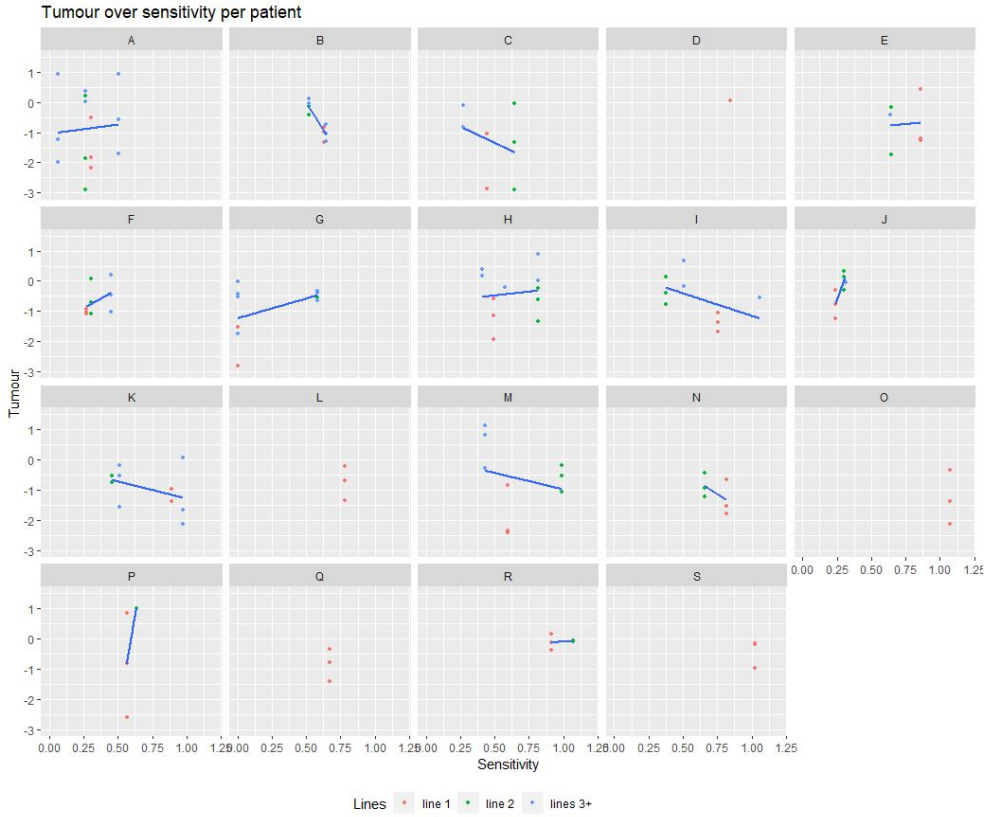


Figure 8: Tumour over sensitivity given a patient

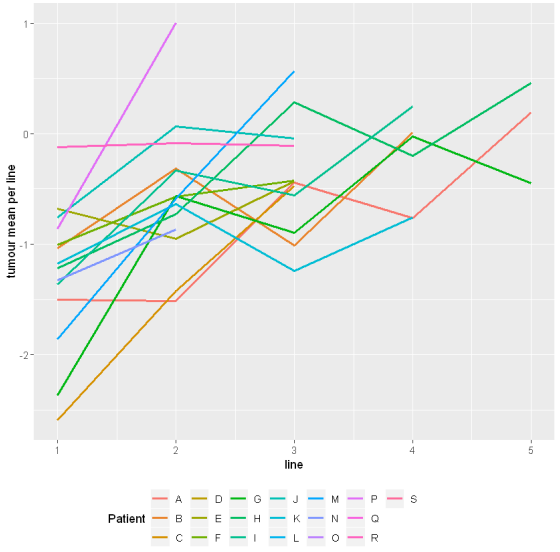


Figure 9: Interaction between patient and line

Plot 9 shows evidence of interaction between the line and patient, i.e. for all the patients the change in tumour marker is less strong in successive treatment lines, but each subject present an individual evolution towards the global trend.

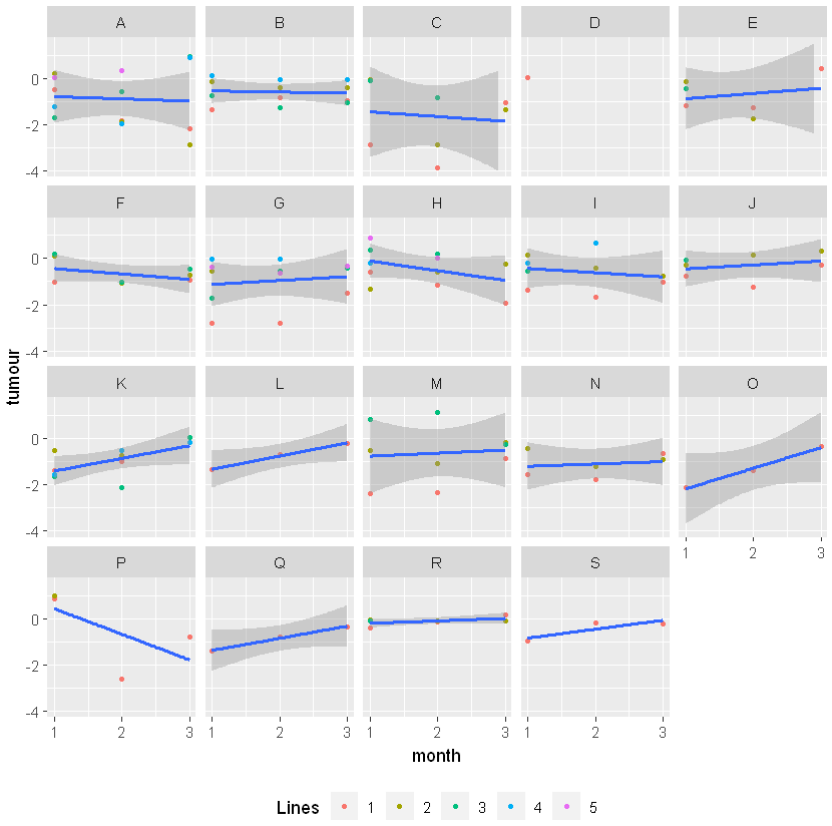


Figure 10: Tumour per month given a patient

A.3 Model building and hypotheses

A.3.1 First model with all the possible interactions

```
mod1.ml = lme(fixed = tumour~sensitivity*line + month*line, random=list(patient=pdBlocked(list(pdSymm(~
    sensitivity),pdIdent(~month-1),pdIdent(~line-1)))), data = df, method = "ML")

Linear mixed-effects model fit by maximum likelihood
 Data: df
       AIC      BIC    logLik
  353.0123 414.3319 -155.5062

Random effects:
 Composite Structure: Blocked

 Block 1: (Intercept), sensitivity
 Formula: ~sensitivity | patient
 Structure: General positive-definite
            StdDev    Corr
(Intercept) 0.4241631 (Intr)
sensitivity 0.1574855 -1

 Block 2: month
 Formula: ~month - 1 | patient
             month
StdDev: 3.0749e-07

 Block 3: line1, line2, line3, line4, line5
 Formula: ~line - 1 | patient
 Structure: Multiple of an Identity
              line1        line2        line3        line4        line5  Residual
StdDev: 5.005847e-05 5.005847e-05 5.005847e-05 5.005847e-05 5.005847e-05 0.7078535

Fixed effects: tumour ~ sensitivity * line + month * line
                       Value Std.Error  DF   t-value p-value
(Intercept)       -2.2392811 0.4087528 104 -5.478326  0.0000
sensitivity        1.0017733 0.4424643 104  2.264077  0.0256
line2              2.1824357 0.5775337 104  3.778889  0.0003
line3              1.9766328 0.5927326 104  3.334780  0.0012
line4              1.5097111 0.7073566 104  2.134300  0.0352
line5              3.0361826 1.1546575 104  2.629509  0.0098
month              0.2348207 0.1256012 104  1.869573  0.0644
sensitivity:line2 -0.8933210 0.6602363 104 -1.353032  0.1790
sensitivity:line3 -1.6190258 0.6264540 104 -2.584429  0.0111
sensitivity:line4 -1.2023342 1.0125517 104 -1.187430  0.2378
sensitivity:line5 -1.4519732 1.5031281 104 -0.965968  0.3363
line2:month       -0.5509093 0.1998203 104 -2.757024  0.0069
line3:month       -0.1014531 0.2264138 104 -0.448087  0.6550
line4:month        0.1517037 0.2916776 104  0.520107  0.6041
line5:month       -0.4305979 0.4278846 104 -1.006341  0.3166
Number of Observations: 137
Number of Groups: 19
```

Listing 3: Summary of model 1

The coefficients of the summary output in Listing (3) correspond to:

- β_0 is the intercept, being the mean tumour change for line 1 and line 1 interaction with sensitivity, line 1 interaction with month
- β_1 is the slope for sensitivity denoted as s
- β_2 is the slope for month denoted as m
- β_3, β_5, β_9, β_{12} is the intercept respectively for line 2, 3, 4 and 5 where line is denoted as L
- β_5, β_8, β_{11}, β_{14} is the slope difference for month between line 1 and line 2, 3, 4 and 5, respectively
- β_4, β_7, β_{10}, β_{13} is the slope difference for sensitivity between line 1 and line 2, 3, 4 and 5, respectively
- $\gamma_{0,i}$ is the random effect due to patient i
- $\gamma_{1,i}$ is the random effect due to interaction between patient i and sensitivity on the mean response
- $\gamma_{2,i}$ is the random effect due to interaction between patient i and month on the mean response
- γ_3 is the random effect due to interaction between patient i and the level j of the factor line on the mean response

A.3.2 Model check of model 1

- Normality assumption for random effects and residuals

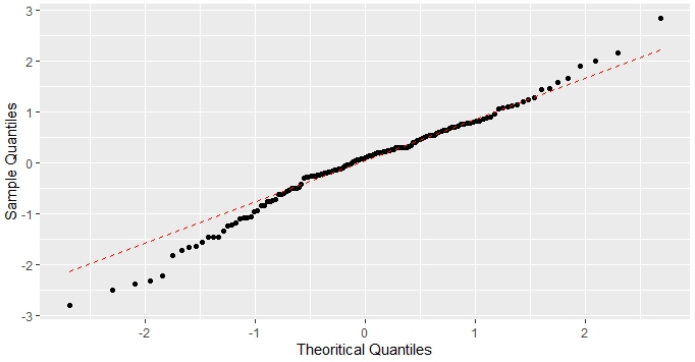


Figure 11: Normality of the residuals model 1

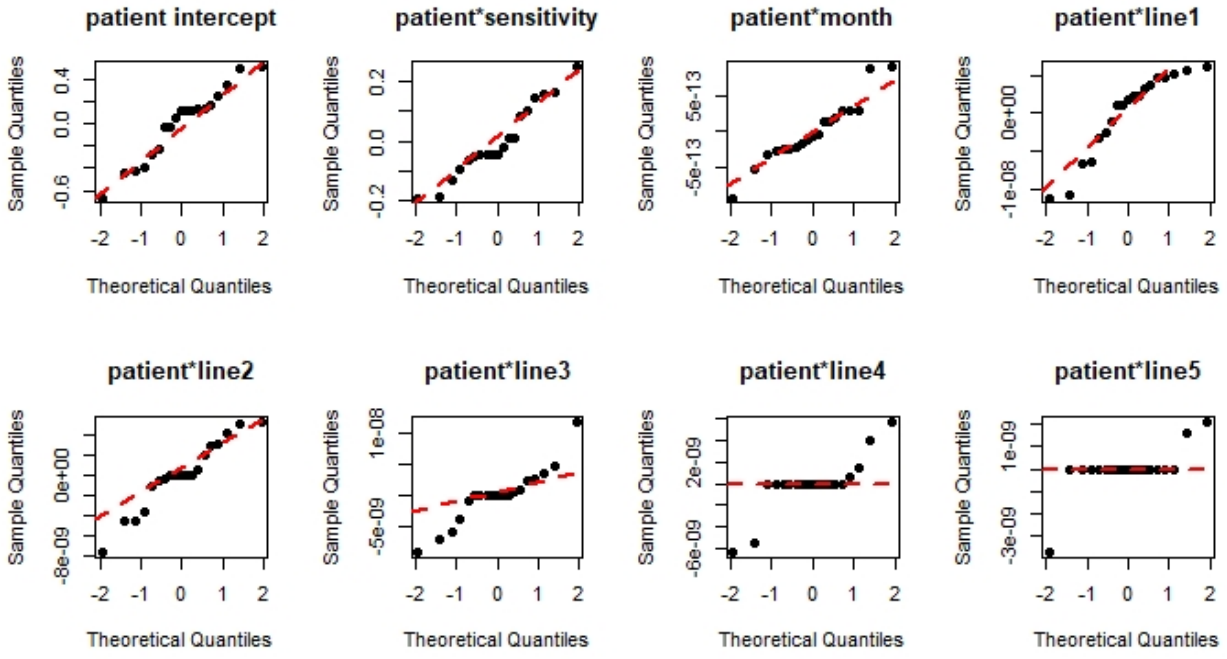


Figure 12: Normality of the random effects

- Homoscedasticity (e.g. equal variance of the response (tumour) around the different line levels) and linearity between the covariates (sensitivity, month) and the response (tumour)

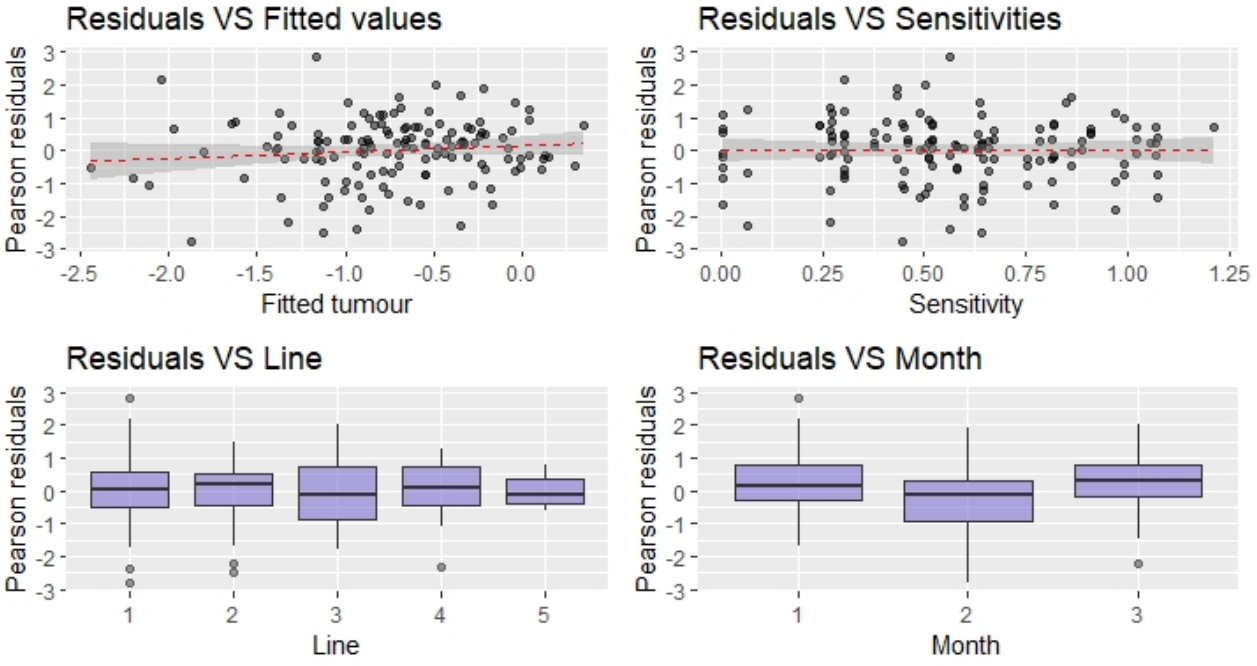


Figure 13: Homoscedasticity and linearity

- Correlation of random effects

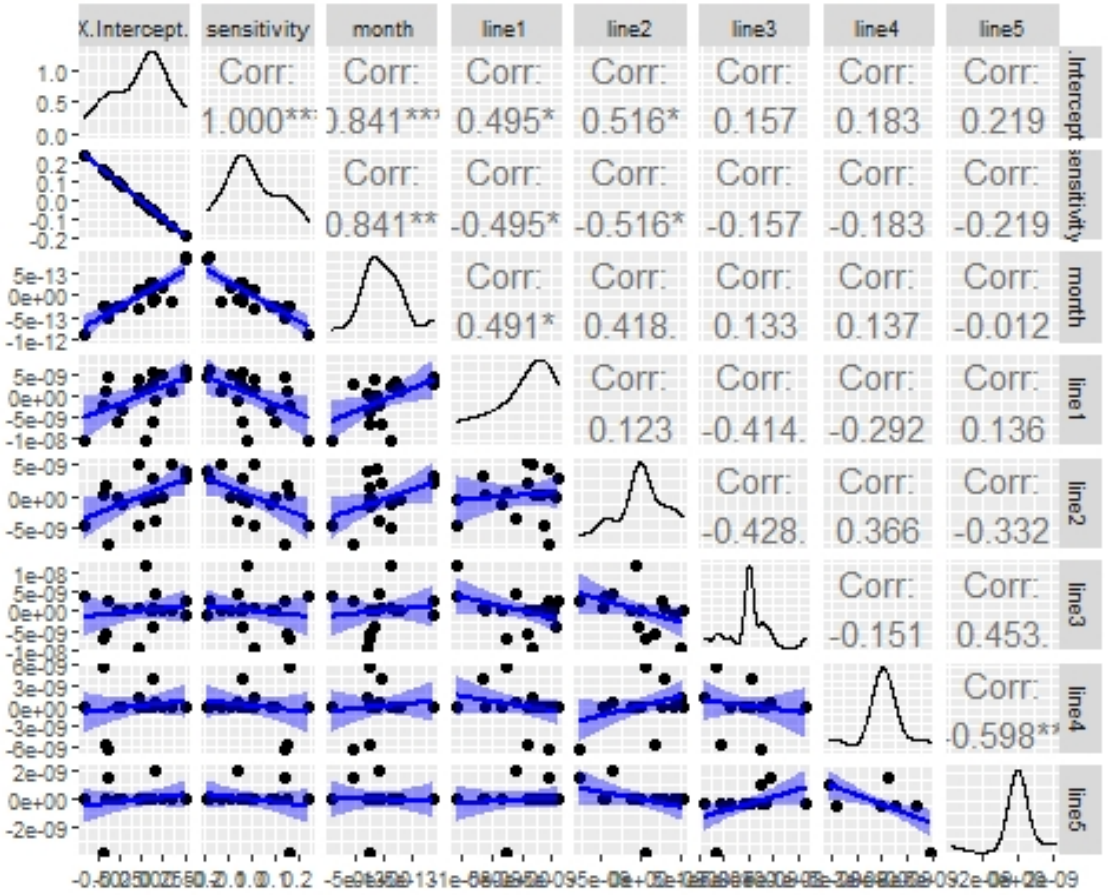


Figure 14: Correlation of random effects

A.3.3 Model building

Since we are testing for the significance of the interaction between fixed effects ($month * line$) and we have unbalanced data, we choose to make an F test with approximate distribution and Kenward-Roger correction for the degrees of freedom.

```
mod2.reml = lmer(tumour~sensitivity*line + month*line+ (1 | patient), data=df, REML=T)
mod3.reml = lmer(tumour~sensitivity*line + month + (1 | patient), data=df, REML=T)

pbkrtest::KRmodcomp(mod2.reml,mod3.reml)

F-test with Kenward-Roger approximation; time: 0.18 sec
large : tumour ~ sensitivity * line + month * line + (1 | patient)
small : tumour ~ sensitivity * line + month + (1 | patient)
          stat      ndf      ddf F.scaling p.value
Ftest   2.4671   4.0000 108.2048   0.99999 0.04913 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC(mod2.reml)
354.68
AIC(mod3.reml)
353.1039
```

Listing 4: Testing for interaction (month*line)

```
> mod5.ml.restr<-lmer(tumour~sensitivity+line + month + I(month^2) +
+                       (1|patient), data = df, REML = F)
> mod5.ml.full<-lmer(tumour~sensitivity*line + month + I(month^2) +
+                      (1|patient), data = df, REML = F)
> n.r = 1000
> mx.t.nr = simulate(mod5.ml.restr, nsim=n.r)
> pval.r = Chisq.r = lrt.r = rep(NA,n.r)
> data.r <- df
> for (rw in 1:n.r){
+   data.r$tumour = mx.t.nr[,rw]
+   fit.restr = lmer(tumour~sensitivity+line + month + I(month^2) + (1|patient), data = data.r, REML =
    FALSE)
+   fit.full = lmer(tumour~sensitivity*line + month + I(month^2) + (1|patient), data = data.r, REML =
    FALSE)
+   lrt.r[rw] = as.numeric(2*(logLik(fit.full) - logLik(fit.restr)))
+   anova.r = anova(fit.restr, fit.full)
+   pval.r[rw] = anova.r$Pr[2]
+   Chisq.r[rw] = anova.r$Chisq[2]
+ }
> LRT.real = as.numeric(2*(logLik(mod5.ml.full) - logLik(mod5.ml.restr)))
> mean(LRT.real<lrt.r) # This is the p-value of LRT corrected with Bootstrap
[1] 0.03
```

Listing 5: LRT corrected with Bootstrap for testing the joint significance between line and sensitivity (without heteroskedasticity)

```
mod4.ml = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, data = df,
    method = "ML")
mod4.ml.hetero = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, weights =
     varIdent(form=~1|line),
                     ,data = df, method = "ML")
anova(mod4.ml, mod4.ml.hetero)

               Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod4.ml            1 14 339.9194 380.7992 -155.9597
mod4.ml.hetero     2 18 343.3558 395.9155 -153.6779 1 vs 2 4.563592  0.3351

mod4.ml_2 = lme(fixed = tumour~sensitivity*line_3 + month +I(month^2), random=~1|patient, data = df,
    method = "ML")
mod4.ml.hetero_2 = lme(fixed = tumour~sensitivity*line_3 + month + I(month^2), random=~1|patient,
    weights=varIdent(form=~1|line_3), data = df, method = "ML")

anova(mod4.ml_2, mod4.ml.hetero_2)
                 Model df      AIC      BIC   logLik   Test   L.Ratio p-value
mod4.ml_2            1 10 336.1840 365.3838 -158.092
mod4.ml.hetero_2     2 12 340.0841 375.1238 -158.042 1 vs 2 0.0999371  0.9513
```

Listing 6: Testing for heteroscedasticity through the different levels of line

```
mod5.ml = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, data = df,
    method = "ML")
mod5.ml.hetero = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, weights=
    varExp(form=~sensitivity),
                     ,data = df, method = "ML")
anova(mod5.ml, mod5.ml.hetero)

 Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5.ml            1 14 339.9194 380.7992 -155.9597
mod5.ml.hetero     2 15 336.8431 380.6428 -153.4215 1 vs 2 5.076333  0.0243

mod5.ml_2 = lme(fixed = tumour~sensitivity*line_3 + month +I(month^2), random=~1|patient, data = df,
    method = "ML")
mod5.ml.hetero_2 = lme(fixed = tumour~sensitivity*line_3 + month + I(month^2), random=~1|patient,
    weights=varExp(form=~sensitivity), data = df, method = "ML")
anova(mod5.ml_2, mod5.ml.hetero_2)
                  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5.ml_2             1 10 336.1840 365.3838 -158.0920
mod5.ml.hetero_2      2 11 333.9589 366.0787 -155.9794 1 vs 2 4.225135  0.0398
```

Listing 7: Testing for heteroscedasticity through sensitivity

A.3.4 Hypotheses

The main questions we are trying to answer are "Is the variable `sensitivity` a good predictor for `tumour`?" and if yes, "Is it a better predictor for `line 1` compared to `line 2` and `lines 3+`?".

For the first question, we use the final model with `line` being a 5 level factor.

$$T_{i(j(m),s)} = \beta_0 + \beta_1 \cdot s + \beta_2 \cdot m + \beta_{15} \cdot m^2 + (\beta_3 + \beta_4 \cdot s) \cdot L_2 + (\beta_6 + \beta_7 \cdot s) \cdot L_3 + (\beta_9 + \beta_{10} \cdot s) \cdot L_4 + (\beta_{12} + \beta_{13} \cdot s) \cdot L_5 + \gamma_{0,i} + \epsilon_{i(j(m),s)}$$

where $\gamma_0 \sim \mathcal{N}(0, \sigma^2_{\gamma_0})$, $\epsilon_{i(j(m),s)} \sim \mathcal{N}(0, \sigma^2_\epsilon exp(2\delta s_i))$. So we regress `sensitivity` on `tumour` and test the coefficient of `sensitivity` as below.

$$H_0 : \beta_1 = 0 \tag{3}$$

$$H_1 : \beta_1 > 0 \tag{4}$$

For the second question, we use the final model with `line` being a 3 level factor.

$$T_{i(j(m),s)} = \beta_0 + \beta_1 \cdot s + \beta_2 \cdot m + \beta_{15} \cdot m^2 + (\beta_3 + \beta_4 \cdot s) \cdot L_2 + (\beta_6 + \beta_7 \cdot s) \cdot L_3 + \gamma_{0,i} + \epsilon_{i(j(m),s)}$$

where $\gamma_0 \sim \mathcal{N}(0, \sigma^2_{\gamma_0})$, $\epsilon_{i(j(m),s)} \sim \mathcal{N}(0, \sigma^2_\epsilon exp(2\delta s_i))$.

We use treatment contrast for `line` with `line 1` as reference and formulate the following hypothesis to test whether the relation between `sensitivity` and `tumour` is positive (the smaller the sensitivity score, the more the tumour shrinks), so we should make a one-sided test for question 2.

$$H_0 : \beta_4 = 0 \quad and \quad \beta_7 = 0 \tag{5}$$

$$H_1 : \beta_4 < 0 \quad and \quad \beta_7 < 0 \tag{6}$$

```
mod5.ml.hetero = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, weights=
    varExp(form=~sensitivity),data = df, method = "ML")

summary(mod5.ml.hetero)

Linear mixed-effects model fit by maximum likelihood
 Data: df
       AIC      BIC    logLik
  336.8431 380.6428 -153.4215

Random effects:
 Formula: ~1 | patient
        (Intercept)  Residual
StdDev:   0.3310928 0.9660155

Variance function:
 Structure: Exponential of variance covariate
 Formula: ~sensitivity
 Parameter estimates:
     expon
-0.5856777
Fixed effects: tumour ~ sensitivity * line + month + I(month^2)
                       Value Std.Error  DF   t-value p-value
(Intercept)       -0.6337419 0.5574104 107 -1.136940  0.2581
sensitivity        1.0444094 0.4438531 107  2.353052  0.0204
line2              1.2665264 0.4498647 107  2.815350  0.0058
line3              1.8965228 0.4292433 107  4.418293  0.0000
line4              1.6823898 0.5764878 107  2.918344  0.0043
line5              2.0054390 0.9713375 107  2.064616  0.0414
month             -1.5891717 0.5191873 107 -3.060883  0.0028
I(month^2)         0.4301362 0.1303458 107  3.299962  0.0013
sensitivity:line2 -1.1059894 0.6382271 107 -1.732909  0.0860
sensitivity:line3 -1.8633186 0.6041921 107 -3.083984  0.0026
sensitivity:line4 -1.0247777 1.1766732 107 -0.870911  0.3858
sensitivity:line5 -0.9521767 1.5204206 107 -0.626259  0.5325

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-2.87714275 -0.62238430  0.09049801  0.64857162  2.47840460

Number of Observations: 137
Number of Groups: 19
```

Listing 8: Final model 1 (5 levels) - Testing for 1st Research Question

```
mod5.ml.hetero_02 = lme(fixed = tumour~sensitivity + line_3 + month + I(month^2), random=~1|patient,
    weights=varExp(form=~sensitivity), data = df, method = "ML")
mod5.ml.hetero_2 = lme(fixed = tumour~sensitivity*line_3 + month + I(month^2), random=~1|patient,
    weights=varExp(form=~sensitivity), data = df, method = "ML")

anova(mod5.ml.hetero_02, mod5.ml.hetero_2)
                  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod5.ml.hetero_02     1  9 339.4960 365.7758 -160.7480
mod5.ml.hetero_2      2 11 333.9589 366.0787 -155.9794 1 vs 2 9.537112  0.0085
```

Listing 9: Testing for 2nd Research Question

```
mod5.ml.hetero_2 = lme(fixed = tumour~sensitivity*line_3 + month + I(month^2), random=~1|patient,
    weights=varExp(form=~sensitivity), data = df, method = "ML")

summary(mod5.ml.hetero_2 )

Linear mixed-effects model fit by maximum likelihood
 Data: df
       AIC      BIC    logLik
  333.9589 366.0787 -155.9794

Random effects:
 Formula: ~1 | patient
        (Intercept)  Residual
StdDev:   0.3174698 0.9598888

Variance function:
 Structure: Exponential of variance covariate
 Formula: ~sensitivity
 Parameter estimates:
     expon
-0.5311948
Fixed effects: tumour ~ sensitivity * line_3 + month + I(month^2)
                          Value Std.Error  DF   t-value p-value
(Intercept)          -0.7313494 0.5543476 111 -1.319298  0.1898
sensitivity           1.1076095 0.4350972 111  2.545660  0.0123
line_32               1.2896066 0.4470256 111  2.884861  0.0047
line_33+              1.9255354 0.3760884 111  5.119901  0.0000
month                -1.5164243 0.5239944 111 -2.893970  0.0046
I(month^2)            0.4111617 0.1315921 111  3.124516  0.0023
sensitivity:line_32  -1.1438099 0.6368013 111 -1.796180  0.0752
sensitivity:line_33+ -1.6977549 0.5537213 111 -3.066082  0.0027

Standardized Within-Group Residuals:
       Min          Q1         Med          Q3         Max
-2.8304543 -0.5899631  0.1220770  0.6365350  2.4604094

Number of Observations: 137
Number of Groups: 19
```

Listing 10: Final model 2 (3 levels)

```
MuMIn::r.squaredGLMM(mod5.ml.hetero)
           R2m       R2c
[1,] 0.2056335 0.2891391
```

Listing 11: Goodness of fit measure for final model 1 (5 levels)

The percentage of variance explained by the fixed part of the final model 1 (with 5 levels) is 20.5%. The percentage of variance explained by the fixed and random parts of this model is 28.91%.

```
MuMIn::r.squaredGLMM(mod5.ml.hetero_2)
           R2m     R2c
[1,] 0.1971149 0.27628
```

Listing 12: Goodness of fit measure for final model 2 (3 levels)

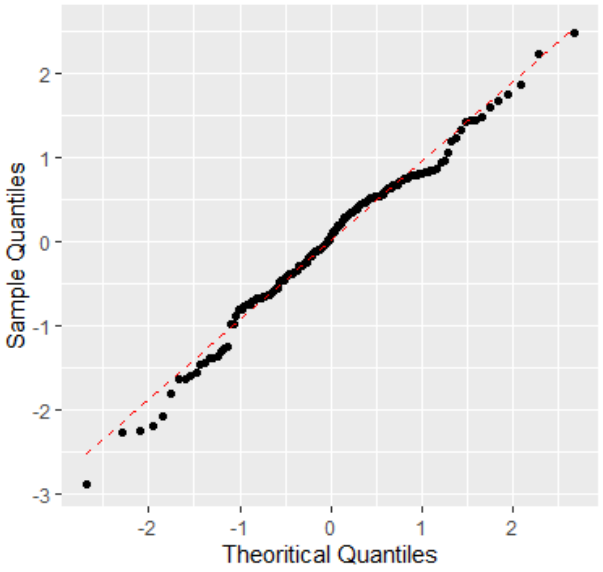
The percentage of variance explained by the fixed part of the final model 2 (with 3 levels) is 19.71%. The percentage of variance explained by the fixed and random parts of this model is 27.62%.

For both models we observe that adding the random intercept increases the explanatory power.

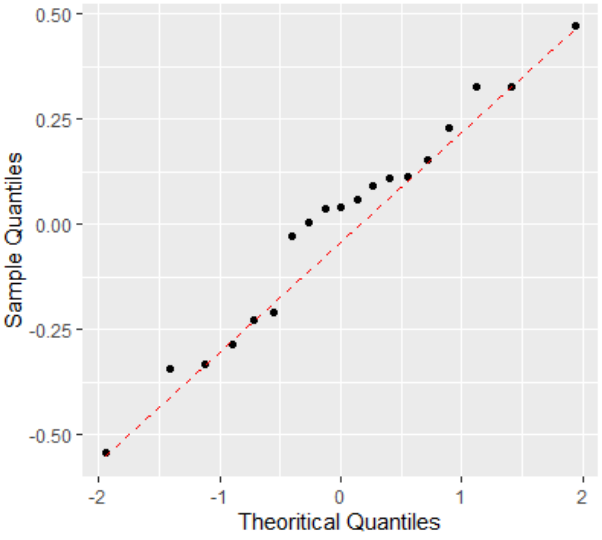
A.4 Model estimation and model check

A.4.1 Model check for final model 1 (line has 5 levels)

- Normality assumption for random effects and residuals



(a) Normality of the residuals



(b) Normality of the random effect

Figure 15: Normality assumption for final model 1

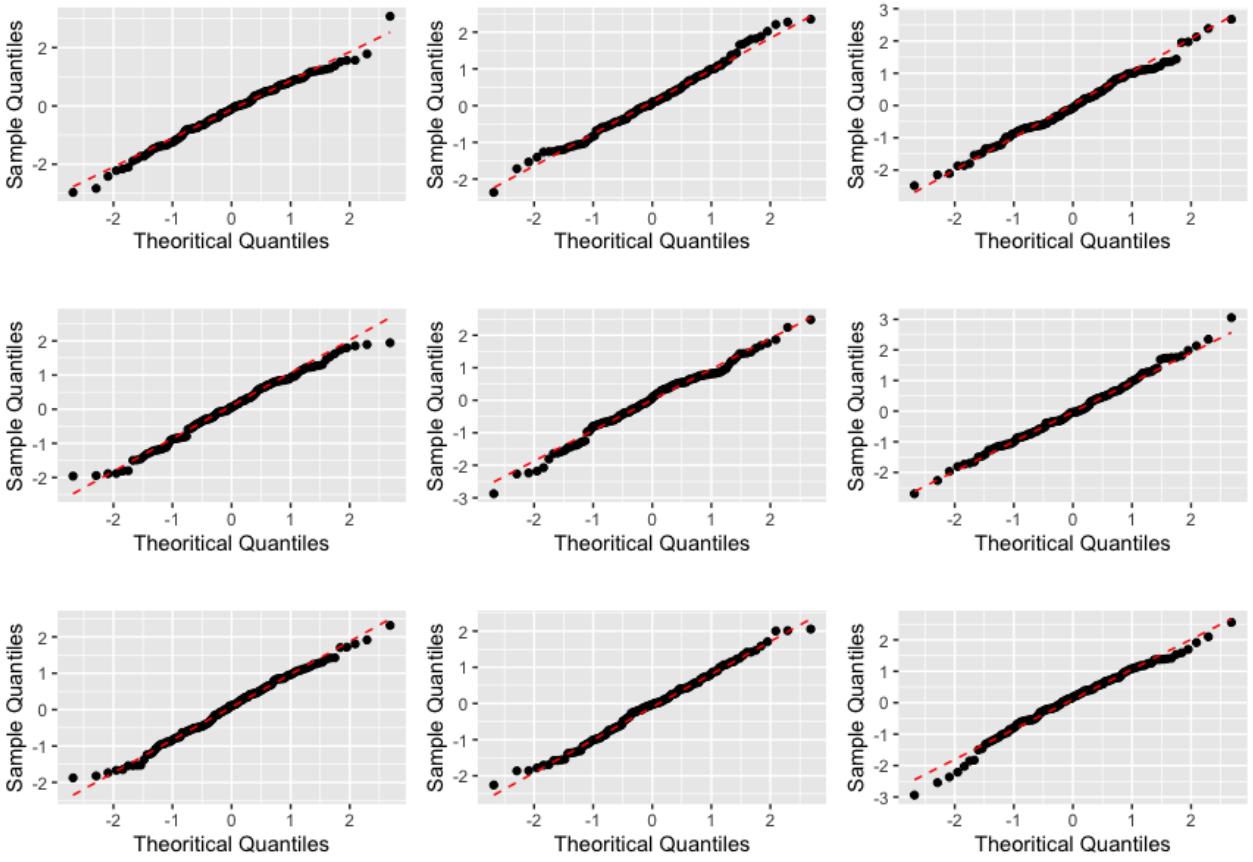


Figure 16: Final model (without heteroskedasticity) residuals vs simulated data $\mathcal{N}(0, \sigma_\epsilon^2)$

- Homoscedasticity

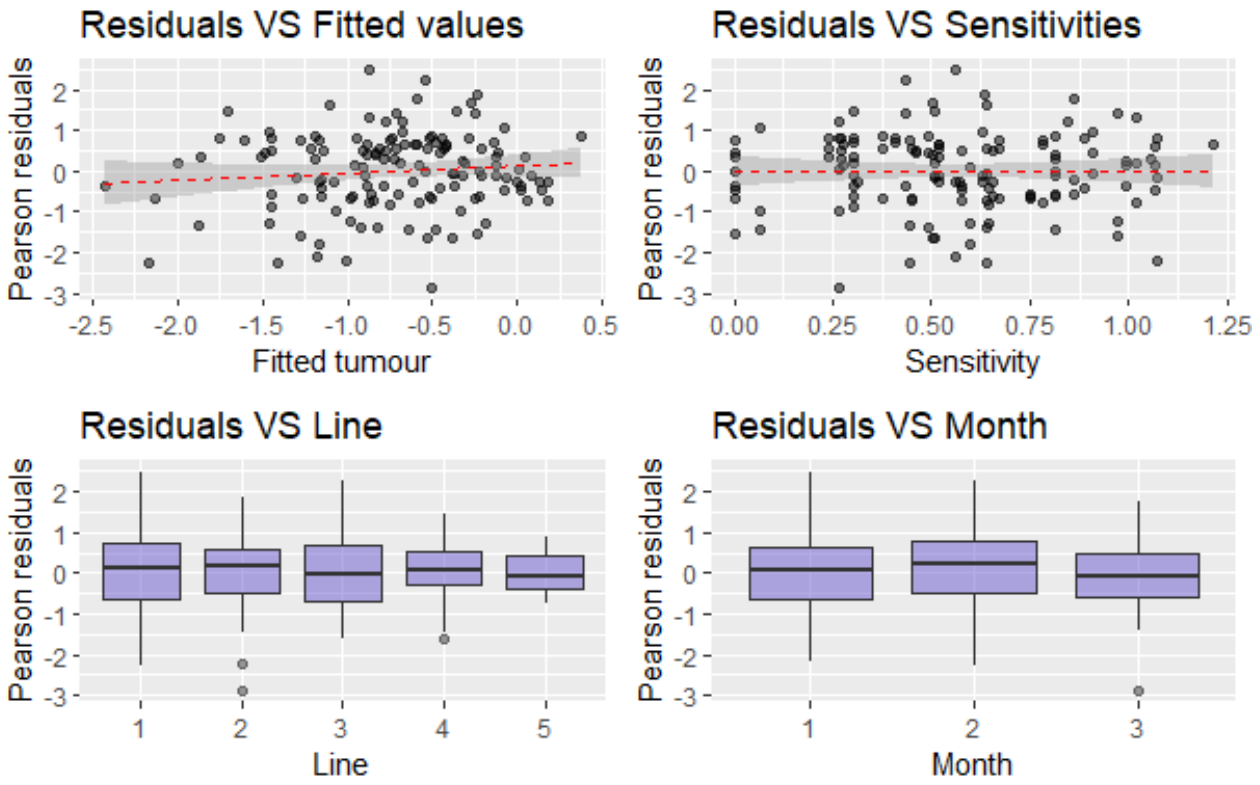
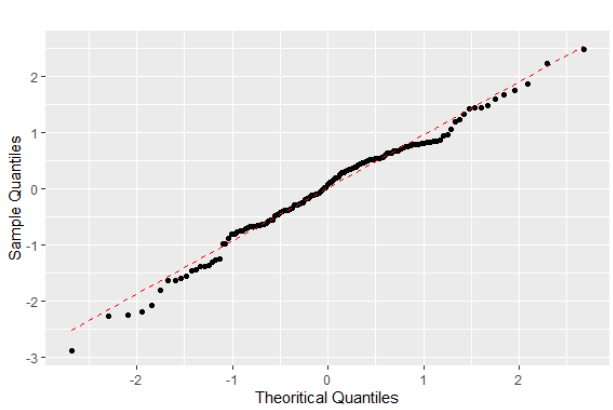


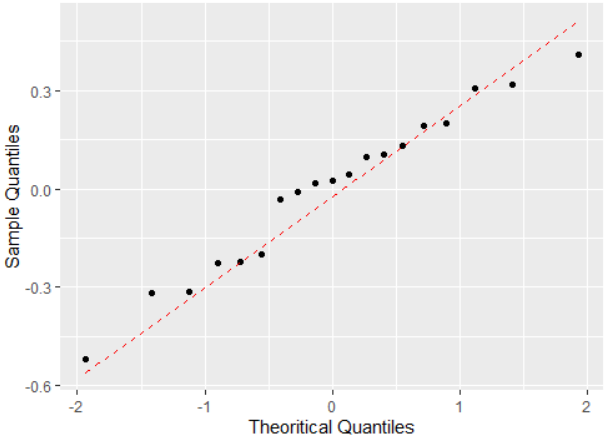
Figure 17: Homoscedasticity and linearity in the full model (5 levels of line)

A.4.2 Model check for final model with 3 levels

- Normality assumption for random effects and residuals



(a) Normality of the residuals (3 levels of line)



(b) Normality of the random effect (3 levels of line)

Figure 18: Normality assumption for final model 2

- Homoscedasticity

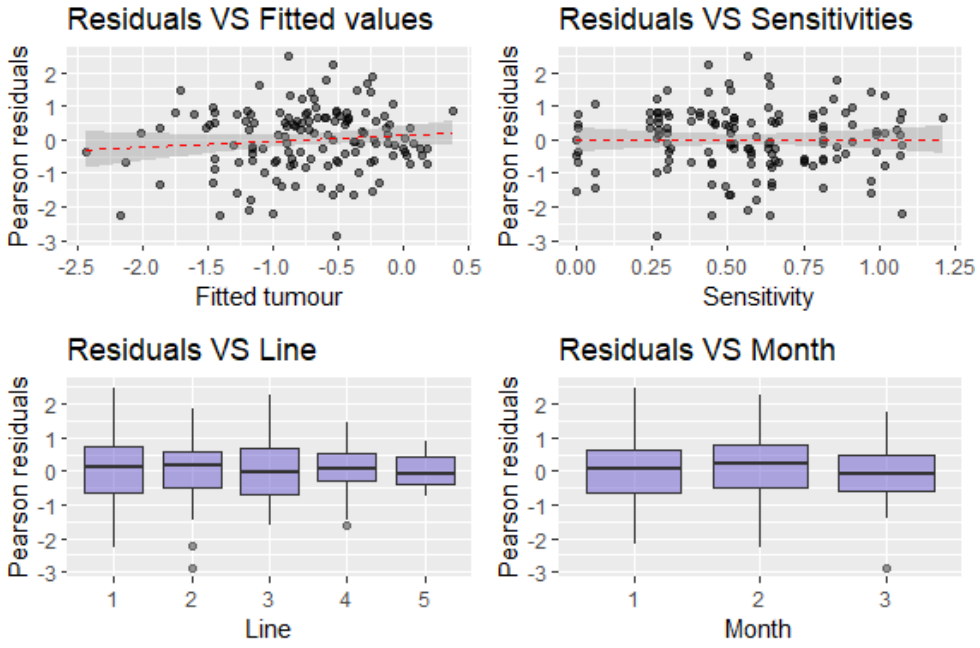


Figure 19: Homoscedasticity and linearity in the full model (3 levels of line)

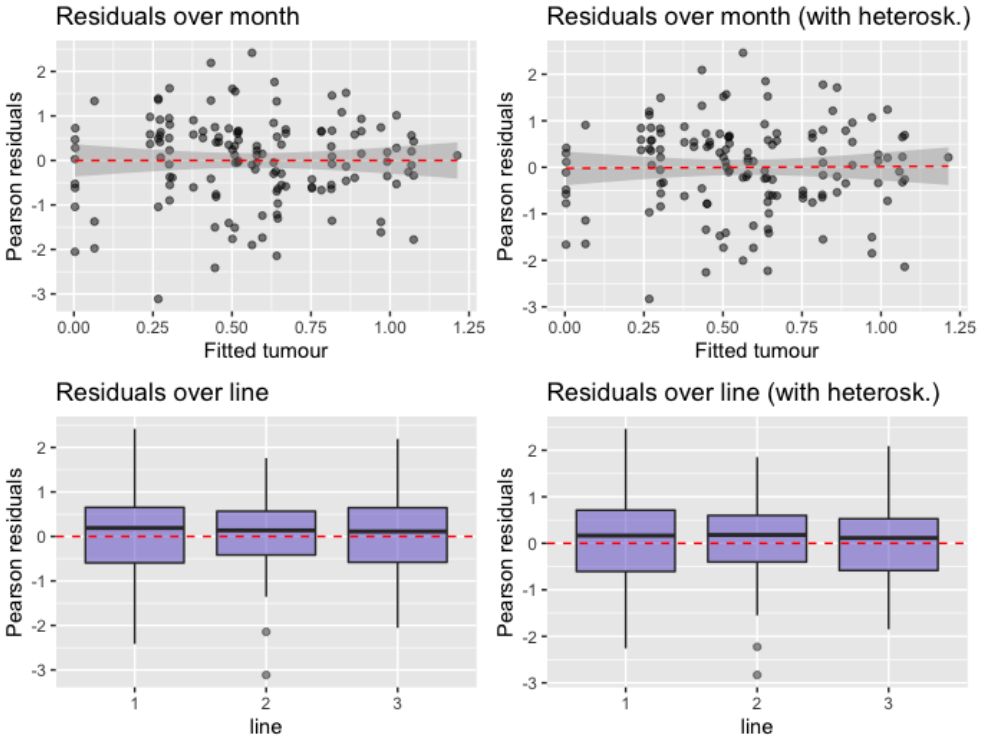


Figure 20: Residual without and with modelling for heteroscedasticity

In Figure 20 we make a comparison between the model check plots of the model that does not account for the heteroscedasticity and the model that accounts for it. They are almost identical. However when testing for the best model, there is statistical evidence in favor of modelling the heteroscedasticity.

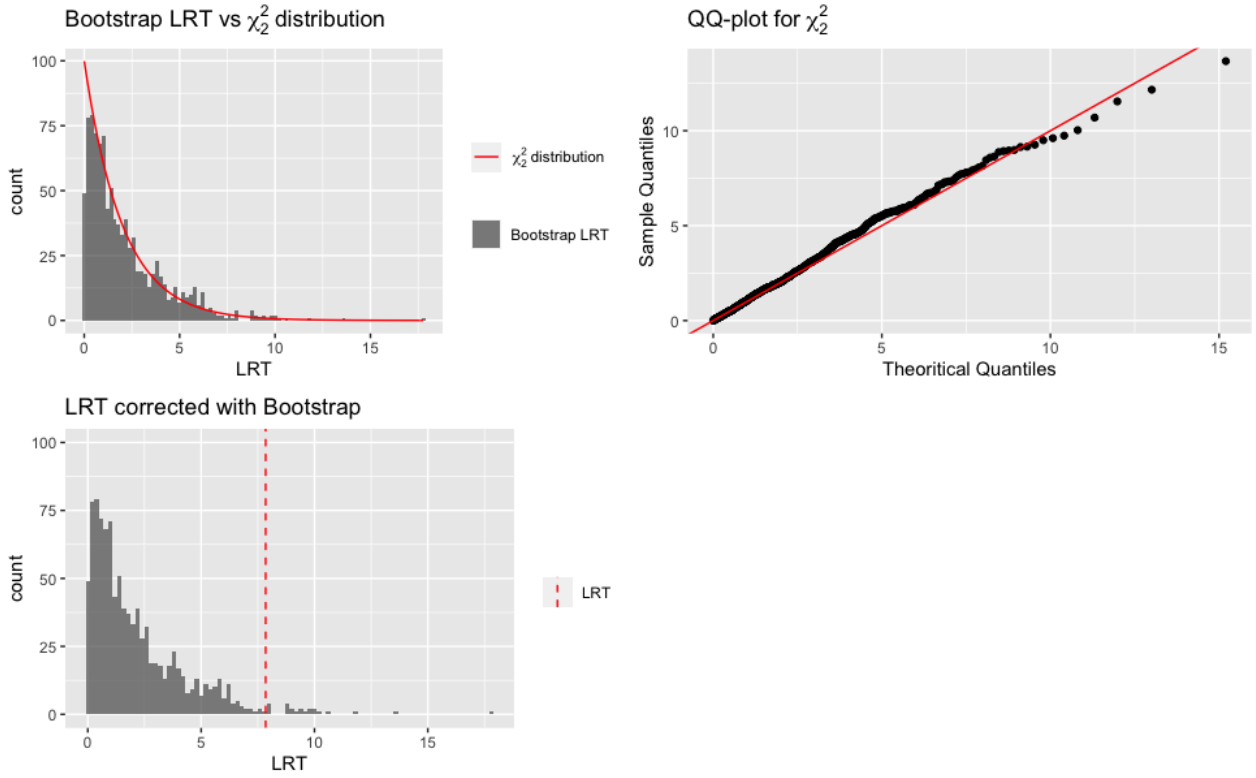


Figure 21: LRT corrected with Bootstrap

In Figure 21, we can see that the LRT distribution follows almost exactly the χ^2_2 one (both pictures on the top). It means that the simple LRT could be quite reliable as well.

A.4.3 Variance-covariance matrix

```
> Omega.matrix = function(model) {
+   if(class(model)[1]=="lmerModLmerTest"|class(model)[1]=="lmerMod"){
+     sigma2.epsilon = sigma(model)^2
+     Psi.star       = crossprod(getME(model,"Lambdat"))*sigma2.epsilon
+     Z              = getME(model,"Z")
+     Omega          = Z %*% Psi.star %*% t(Z) + sigma2.epsilon* Diagonal(nrow(Z))
+     Omega
+   }else{
+     warning("Function only works on outcput of function lmer()\n")
+   }
+ }
> Omega.plot = function(Omega,legend=TRUE,axes=TRUE){
+   corw = cov2cor(Omega)
+   if(any(corw<0)){
+     colw=c(gray(1),rainbow(197)[197:100],gray(.9),rainbow(197)[99:1],gray(0))
+     zlim=c(-1,1)
+   }else{
+     colw=c(gray(.9),rainbow(98)[98:1],gray(0))
+     zlim=c(0,1)
+   }
+   image(z=as.matrix(corw[nrow(corw):1,]),zlim=zlim,axes=FALSE,col=colw)
+   if(legend){
+     valw = as.numeric(names(table(as.matrix(corw))))
+     posw = round(valw*length(colw))
+     posw[posw==0] = 1
+     posw[posw>length(colw)] = length(colw)
+     legend("topright",ncol=1,legend=format(round(valw,4)),
+            col=colw[posw],pch=15,bg="light gray",
+            title="Values",box.lwd=NA)
+   }
+   if(axes){
+     axis(3,at=seq(0,1,length=nrow(Omega)),labels=FALSE)
+     axis(2,at=seq(0,1,length=nrow(Omega)),labels=FALSE)
+     axis(2,at=c(1,0),c(1,nrow(Omega)),las=2)
+     axis(3,at=c(0,1),c(1,nrow(Omega)),las=1)
+ }}
```

Listing 13: Code for the omega function

```
> mod5.ml<-lmer(tumour~sensitivity+line + month + I(month^2) +
+                        (1|patient), data = df, REML = F)
> mod5.ml.sum = summary(mod5.ml)
> mod5.ml.sum$varcor
 Groups    Name        Std.Dev.
 patient   (Intercept) 0.30141
 Residual              0.74013
```

Listing 14: Standard deviations of the random intercepts and residuals (without heteroscetasticity)

The `omega()` function does not allow for any model from `lme()` so, again, we will only consider `mod5.ml` (which is equivalent to our best model but without heteroskedasticty because we can not model it with `lmer()`) to build the variance-covariance matrix. Since we only have one random effect which is not correlated with any other parameter, the ICC (Intra-Cluster Correlation) is only about the value of exactly 1 on the diagonal and one value out of the diagonal, which can be calculated with the following formula:

$$ICC = \frac{\sigma^2_{\gamma_0}}{\sigma^2_{\gamma_0} + \sigma^2_{\epsilon}} = 0.1422$$

with $\sigma_{\gamma_0} = 0.3014$ and $\sigma_{\epsilon} = 0.7401$.

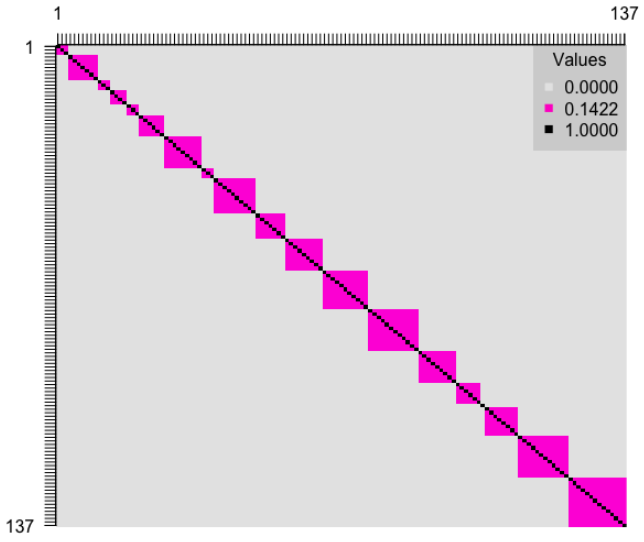


Figure 22: Variance-covariance matrix

```
> mod5.ml.hetero = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, data =
    df, weights=varExp(form=~sensitivity), method = "ML")
summary(mod5.ml.hetero)
...
Random effects:
 Formula: ~1 | patient
        (Intercept)  Residual
StdDev:   0.3310928 0.9660155
```

Listing 15: Standard deviations of the random intercepts and residuals (with heterosketasticity)

Even if we can not use `omega()` to build the variance-covariance matrix, we can still compute the ICC:

$$ICC = \frac{\sigma^2_{\gamma_0}}{\sigma^2_{\gamma_0} + \sigma^2_{\epsilon}} = 0.1051$$

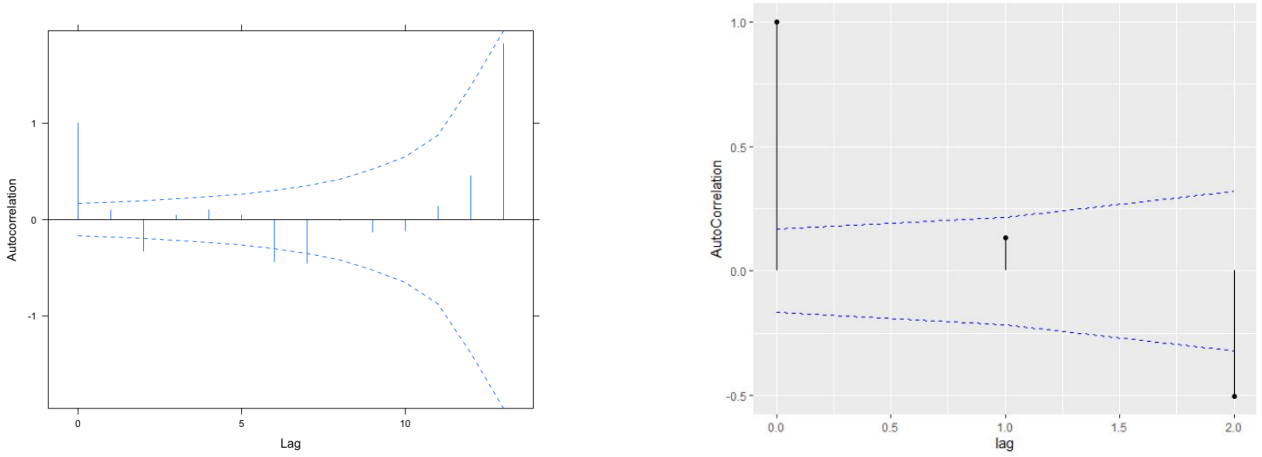
with $\sigma_{\gamma_0} = 0.3311$ and $\sigma_{\epsilon} = 0.9660$

Considering the heteroscedasticity in the model has changed the ICC (from 0.1422 to 0.1051).

A.5 Additional remarks

We implement models that consider correlated residuals, between patient correlation(we assumed no serial correlation between the observations of a patient previously), and robustness. The aim is to increase the power of the model.

A.5.1 ARMA



(a) ACF at the `patient` level

(b) ACF at the `line` within `patient` level

Figure 23: Auto-correlation functions

We consider the ARMA model for modelling correlated residuals. We can observe from Figure 23 that some lags are above the threshold. The issue with this plot is that it is modelling the correlation of the residuals between lines for a patient and not within a line for a patient. The number of lags in the plot indicates that the ACF considers correlation between lines for one patient (max.lag=14-1=13). We know that within each patient within each line, the `tumour` biopsy is examined from month to month. However, we do not know the time lag between treatment lines, and it might not be evenly spaced. This violates the required condition for using ACF plot which needs evenly spaced measurements. To model the correct residual correlation within patient within line, we wrote a function plotting the ACF. See Listing 18. In our case the maximum number of observation is 3 with respect to 3 months or less examination for each treatment.(max.lag=3-1=2). See Figure 23(b).

We try to model autocorrelation with an ARMA(2,1), but the `correlation` argument of lme() considers ARMA within patient and not within patient within line. The argument above still holds since we cannot use ARMA for unevenly spaced lags.

```
> mod5.ml = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, data = df,
    method = "ML")
> mod11.ml = lme(fixed = tumour~sensitivity*line + month +
+                  I(month^2), random=~1|patient, data = df, weights=varExp(form=~sensitivity), corr=
    corARMA(form=~1|patient,p=2,q=1), method = "ML")
> anova(mod5.ml, mod11.ml)
         Model df      AIC      BIC    logLik   Test L.Ratio p-value
mod5.ml      1 10 336.1840 365.3838 -158.0920
mod11.ml     2 14 322.9585 363.8382 -147.4793 1 vs 2 21.2255   3e-04
```

Listing 16: Modelling heteroscedasticity and autocorrelation within `patient`

Consequently, we attempt to solve this problem by using the form $= \sim 1/\text{patient}/\text{line}$. We consider ARMA(0,2) because of the significant correlation for lag 2. See Figure 23. Although we find better results we do not use them for our final model because we do are not certain about the underlying code for the `correlation` argument of `lme()` and we are concerned about the fact that maybe it considers `line` as a random effect, so we do not add it to our model.

```
> mod5.ml = lme(fixed = tumour~sensitivity*line + month + I(month^2), random=~1|patient, data = df,
    method = "ML")
> mod12.ml = lme(fixed = tumour~sensitivity*line + month +
+                   +I(month^2), random=~1|patient, data = df, weights=varExp(form=~sensitivity), corr
    =corARMA(form=~1|patient/line,p=0,q=2), method = "ML")
> anova(mod5.ml, mod12.ml)
         Model df      AIC      BIC    logLik   Test L.Ratio p-value
mod5.ml      1 10 336.1840 365.3838 -158.0920
mod12.ml     2 13 330.6950 380.3347 -148.3475 1 vs 2 15.2244   0.0016
```

Listing 17: Modelling heteroscedasticity and autocorrelation within `patient` within `line`

```r
ACF_nested_structure <- function(res.vec = residuals_vector , df=original_dataframe , o.e=outer.effect , n.e = nested.effect){

  require(ggplot2)
  require(dplyr)

  ####residual dataframe

  res.df = cbind(df,res.vec)

  ####residual mean value

  mean_res=mean(res.vec)

  #### finding maximum lag

  res.df %>%
    group_by_( o.e , n.e ) %>%
    summarise(num=n()) -> max.lag.df

  max.lag<-max(max.lag.df$num)

  #### splitting dataframe to newlist each element of newlist include different nested effect of different outer effect

  res.df %>%
    group_by_( o.e , n.e ) %>%
    group_split(.) -> newlist

  #### temporary list

  temp_list<-newlist

  ####result_dataframe

  result_dataframe<-data.frame(matrix(data = 0,nrow = max.lag,ncol = 6))
  colnames(result_dataframe)<-c("lag","added.values.sum","AutoCovariance","AutoCorrelation","lag.counter","half.confidence.interval")

  result_dataframe$lag<-as.numeric(0:(max.lag-1))

  ####
  for (k in 1:max.lag){

    for (i in 1:length(temp_list)) {

      if(dim(temp_list[[i]])[1]>(k-1)){

        for (j in k:dim(temp_list[[i]])[1]){

          result_dataframe$lag.counter[k]<-result_dataframe$lag.counter[k]+1 #counter of observations for each lag number

          result_dataframe$added.values.sum[k] <- result_dataframe$added.values.sum[k] + (temp_list[[i]]$res.vec[j-(k-1)]-mean_res)*(temp_list[[i]]$res.vec[j]-mean_res) ## added value for covariance term




        }
      }
    }

    result_dataframe$AutoCovariance[k]  <- result_dataframe$added.values.sum[k] / result_dataframe$lag.counter[k]  #autocrrelation

    result_dataframe$AutoCorrelation[k] <- result_dataframe$AutoCovariance[k] / result_dataframe$AutoCovariance[1] #autocrrelation

    result_dataframe$half.confidence.interval [k] <- 1/sqrt(result_dataframe$lag.counter[k])
                  #confidence interval for Autocorrealtion term

  }

  plot <-ggplot(result_dataframe, aes(x=lag, y=AutoCorrelation)) +
          #plot
    geom_point() +
```

```
    geom_segment(aes(x=lag,xend=lag,y=0,yend=AutoCorrelation))+                                    #
    Autocorrelation value for each lag
    geom_line(aes(x = lag, y = 1.96*lagnumbers), data = acf.df,linetype=2,color="blue")+    #upper
    bound C.I
    geom_line(aes(x = lag, y = -1.96*lagnumbers), data = acf.df,linetype=2,color="blue")     ##lower
    bound C.I


  return(plot)
}


ACF_nested_structure(res.vec = res, df = Chemo_therapy , o.e = "patient", n.e = "line")
```

Listing 18: ACF function code for nested strucure (within `patient` within `line`)

A.5.2 Between cluster correlation

In this section we model the negative correlation between patients (ICC only models a correlation ≥ 0). From Listing 19, we get a p-value larger than 5% therefore we will not take into account this model for further analysis.

```
> mod5.ml = lme(fixed = tumour~sensitivity*line + month +
+                    I(month^2), random=~1|patient, data = df, method = "ML")
> mod5.ml.Cor = lme(fixed = tumour ~ sensitivity*line + month +
+                      I(month^2), random = ~1|patient,correlation = corCompSymm(form = ~ 1|patient),
    data = df, method = "ML")
> anova(mod5.ml,mod5.ml.Cor)
            Model df      AIC      BIC    logLik   Test      L.Ratio p-value
mod5.ml         1 10 336.184 365.3838 -158.092
mod5.ml.Cor     2 11 338.184 370.3038 -158.092 1 vs 2 1.119815e-10       1
```

Listing 19: Modelling autocorrelation

A.5.3 Robustness

From the visual inspection we can still notice some outliers and they are not extreme. Since there are not a lot of observations per cluster, they might affect our assumptions (also the standard errors) and lead to wrong conclusions. Therefore we use a robust model. In Listing 21, 26% of the residuals are underweighted and considered as outliers. These new weights have a range of $[0.38, 0.99]$. The weights of the random effect are the same as in the case of a Maximum Likelihood fit. The estimated parameters are roughly the same thus we can conclude that we do not have model deviations which have a strong effect on our conclusions.

We also try to tune parameters for the robustness. We see different estimations for the robust model with `rlmer()` via tuning parameter $k = 50$ (without robustness) for both random effects, residuals and reml (unbiased estimation of variances). This inconsistency encourages us to just trust the default tuning value $k = 1.345$ for both random effects and residuals. See Figure 27.

```
mod6.rlmer <- rlmer(tumour~sensitivity*line + month + I(month^2) + (1| patient), df)

Robust linear mixed model fit by DAStau
Formula: tumour ~ sensitivity * line + month + I(month^2) + (1 | patient) Data: df

Random effects:
 Groups   Name        Variance Std.Dev.
 patient  (Intercept) 0.1022   0.3197
 Residual             0.5008   0.7076
Number of obs: 137, groups: patient, 19

Fixed effects:
                  Estimate Std. Error t value
(Intercept)        -0.6352     0.5438  -1.168
sensitivity         1.0285     0.4196   2.451
line2               1.3071     0.4157   3.144
line3               1.7504     0.3782   4.628
line4               1.7847     0.4451   4.009
line5               2.1281     0.8673   2.454
month              -1.5087     0.5292  -2.851
I(month^2)          0.4025     0.1330   3.027
sensitivity:line2  -1.1474     0.6445  -1.780
sensitivity:line3  -1.7337     0.6077  -2.853
sensitivity:line4  -1.2809     0.9683  -1.323
sensitivity:line5  -1.2415     1.4580  -0.852

Robustness weights for the residuals:
 109 weights are ~= 1. The remaining 28 ones are summarized as
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.392   0.703   0.795   0.786   0.933   0.999

Robustness weights for the random effects:
 All 19 weights are ~= 1.

Rho functions used for fitting:
  Residuals:
    eff: smoothed Huber (k = 1.345, s = 10)
    sig: smoothed Huber, Proposal II (k = 1.345, s = 10)
  Random Effects, variance component 1 (patient):
    eff: smoothed Huber (k = 1.345, s = 10)
    vcp: smoothed Huber, Proposal II (k = 1.345, s = 10)
```

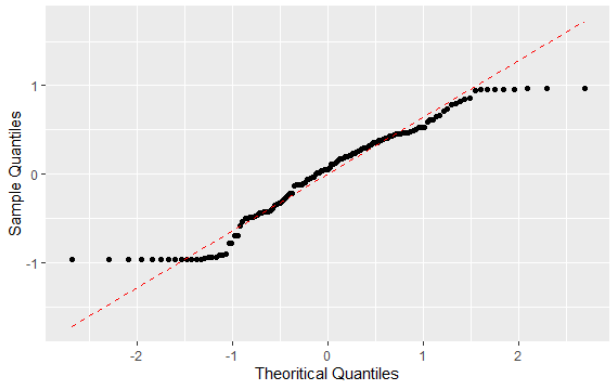
Listing 20: Robust model (line with 5 levels)

```
mod6.rlmer <- rlmer(tumour~sensitivity*line_3 + month + I(month^2) + (1| patient), df)
Robust linear mixed model fit by DAStau
Formula: tumour ~ sensitivity * line_3 + month + I(month^2) + (1 | patient) Data: df

Random effects:
 Groups   Name        Variance Std.Dev.
 patient  (Intercept) 0.07704  0.2776
 Residual             0.51751  0.7194
Number of obs: 137, groups: patient, 19

Fixed effects:
                    Estimate Std. Error t value
(Intercept)          -0.7375     0.5423  -1.360
sensitivity           1.1094     0.4083   2.717
line_32               1.3397     0.4163   3.218
line_33+              1.8220     0.3310   5.504
month                -1.4493     0.5364  -2.702
I(month^2)            0.3878     0.1348   2.877
sensitivity:line_32  -1.1931     0.6428  -1.856
sensitivity:line_33+ -1.5686     0.5416  -2.896

Robustness weights for the residuals:
 108 weights are ~= 1. The remaining 29 ones are summarized as
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.389   0.702   0.834   0.785   0.915   0.996

Robustness weights for the random effects:
 All 19 weights are ~= 1.

Rho functions used for fitting:
  Residuals:
    eff: smoothed Huber (k = 1.345, s = 10)
    sig: smoothed Huber, Proposal II (k = 1.345, s = 10)
  Random Effects, variance component 1 (patient):
    eff: smoothed Huber (k = 1.345, s = 10)
    vcp: smoothed Huber, Proposal II (k = 1.345, s = 10)
```

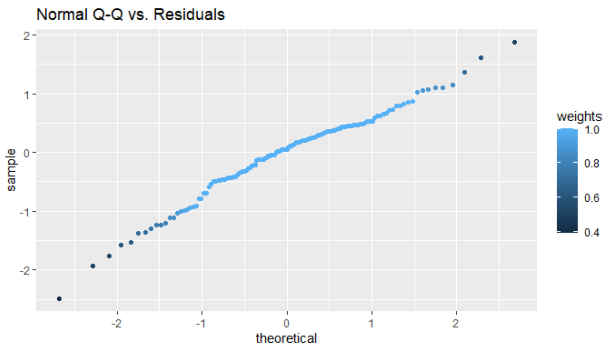
Listing 21: Robust model (line with 3 levels)

A.5.4 Model check for robust model with 3 levels

- Normality assumption for random effects and residuals

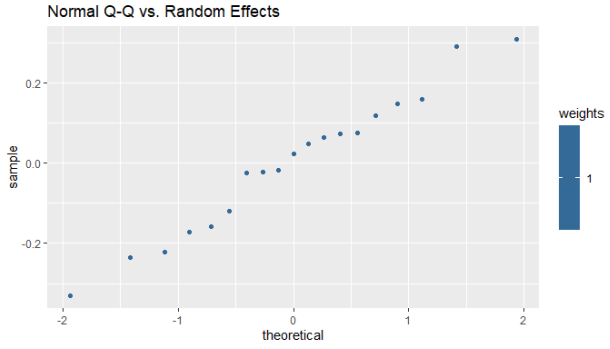


(a) Normality of the residuals

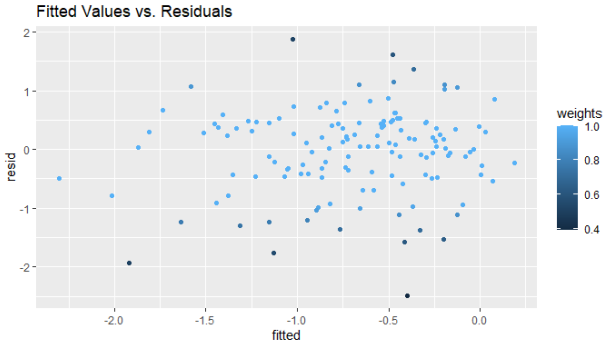


(b) Normality of the residuals

Figure 24: Normality of the residuals of the robust model



(a) Normality of the random effect



(b) Weights of the fitted values

Figure 25: Model check for robust Model

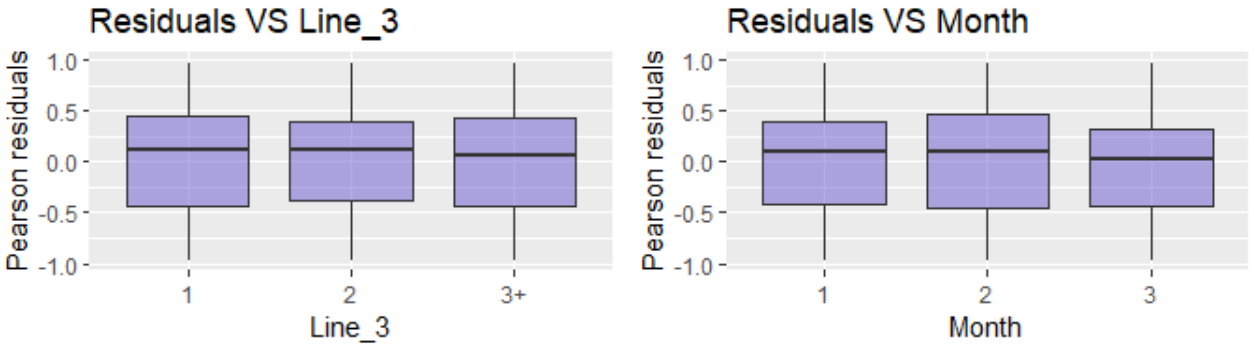
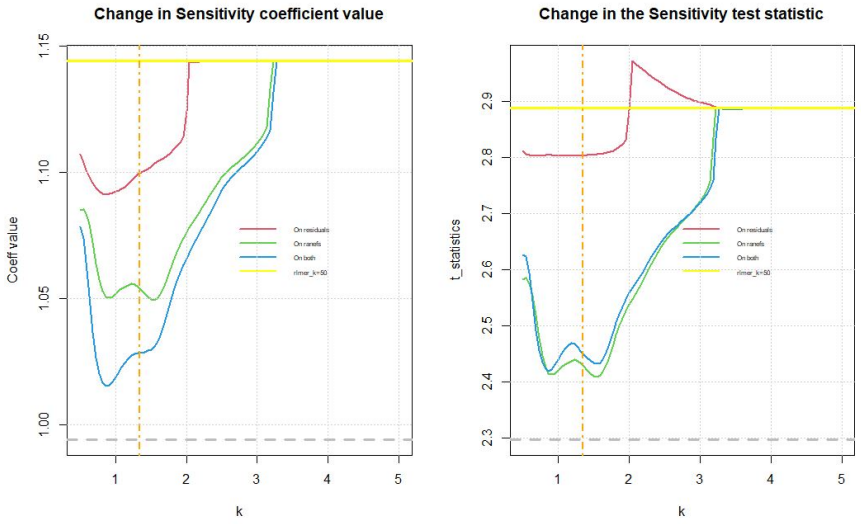
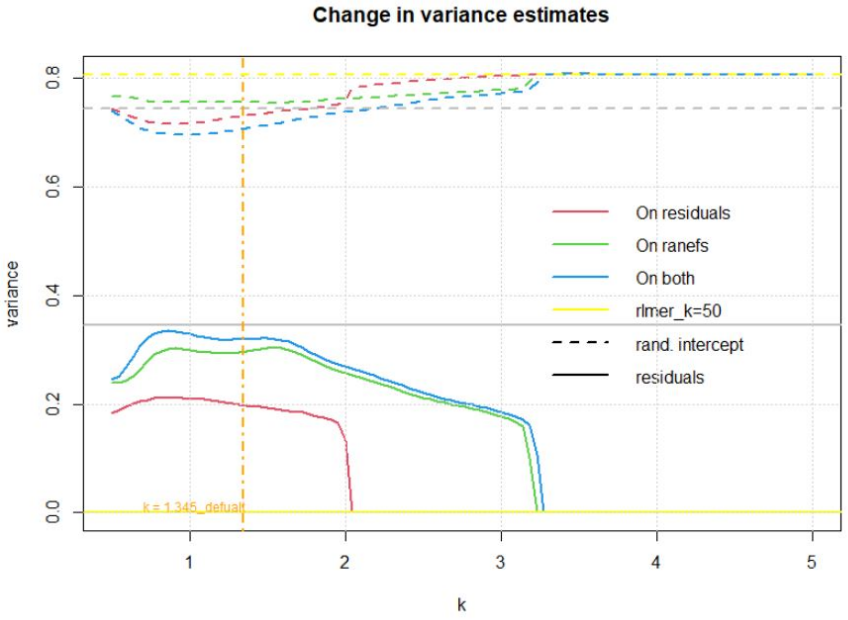


Figure 26: Normality of the residuals for the robust model



(a) Estimation of sensitivity



(b) Estimation of variance for random effect and residual

Figure 27: Robustness parameter tuning