

Q2

Hun Kang

2024-10-27

Q2.

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.3.3
```

```
dat = dget("https://www2.stat.duke.edu/~pdh10/Teaching/610/Homework/Earthquake")
head(dat)
```

```
## Grouped Data: accel ~ distance | Quake
```

```
## <environment: 0x000001c9239bb6f0>
```

```
##      Quake Richter distance soil accel
```

```
## 132    20        5        7.5    1 0.264
```

```
## 133    20        5        8.8    1 0.263
```

```
## 134    20        5        8.9    1 0.230
```

```
## 135    20        5        9.4    1 0.147
```

```
## 136    20        5        9.7    1 0.286
```

```
## 137    20        5        9.7    1 0.157
```

```
# response: accel
```

```
# group: Quake
```

```
# micro level: distance, soil -> fixed and random
```

```
table(dat$soil, dat$Quake)
```

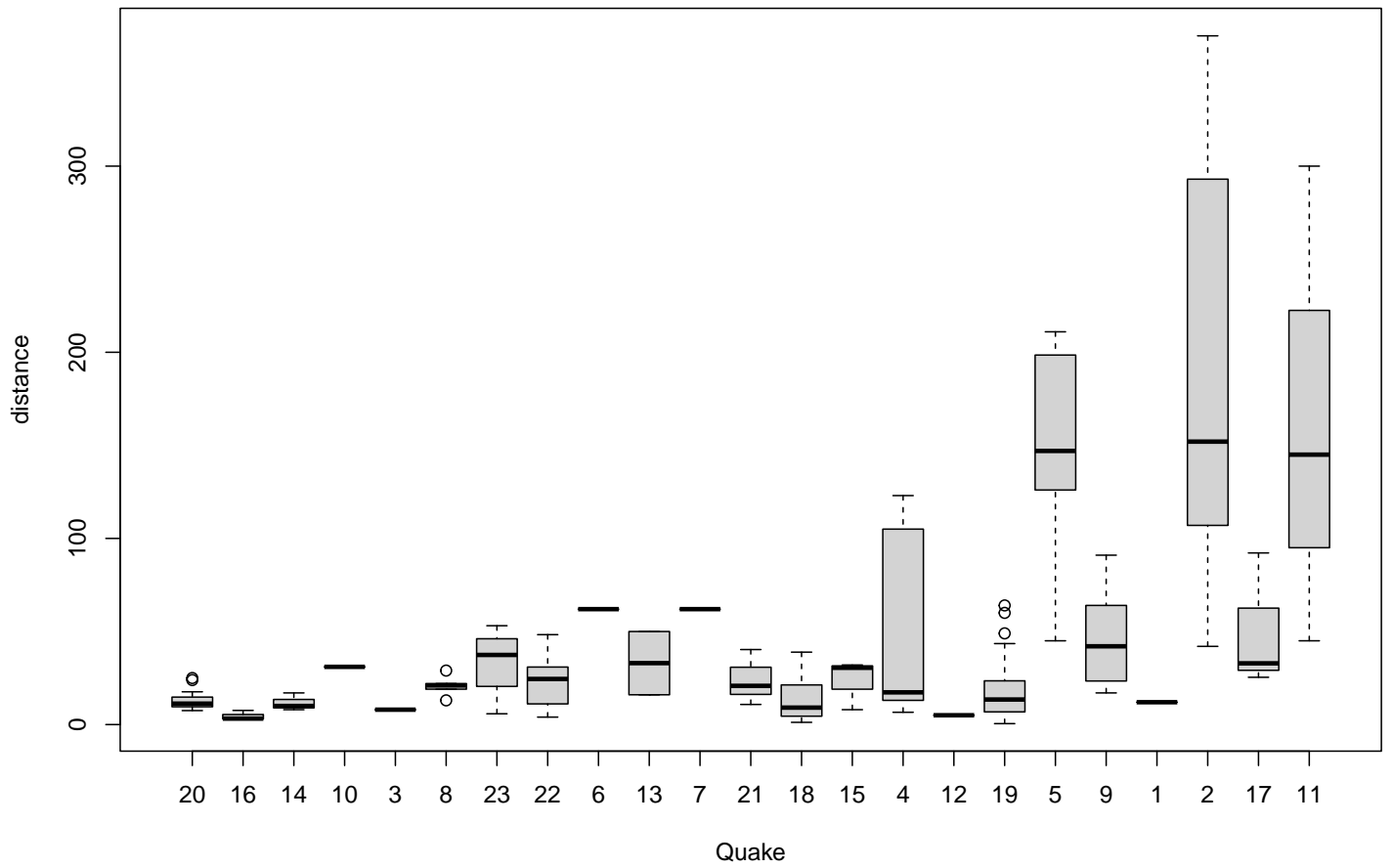
```
##
```

```
##      20 16 14 10  3  8 23 22  6 13  7 21 18 15  4 12 19  5  9  1  2 17 11
```

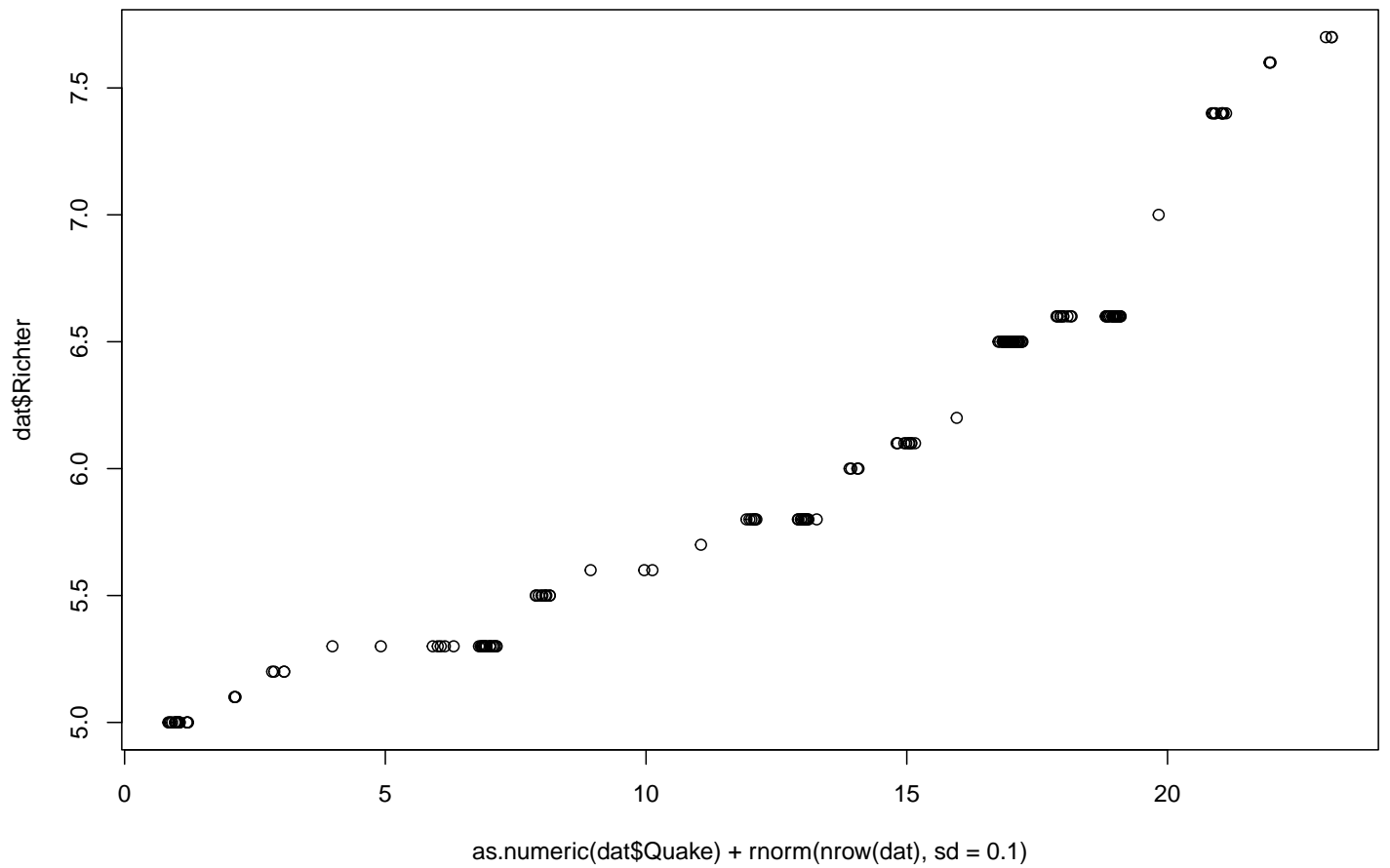
```
##      0  0  0  1  0  1  2  0  0  0  1  0  0  3  2  2  0  2  4  8  0  1  0  2
```

```
##      1 16  3  3  1  0  3 18 10  1  1  1  7  8  2  7  1 36  7 14  1  9  3  1
```

```
boxplot(distance ~ Quake, dat=dat)
```



```
# macro level: Richter -> fixed only
plot(as.numeric(dat$Quake)+rnorm(nrow(dat),sd=0.1), dat$Richter)
```



a.

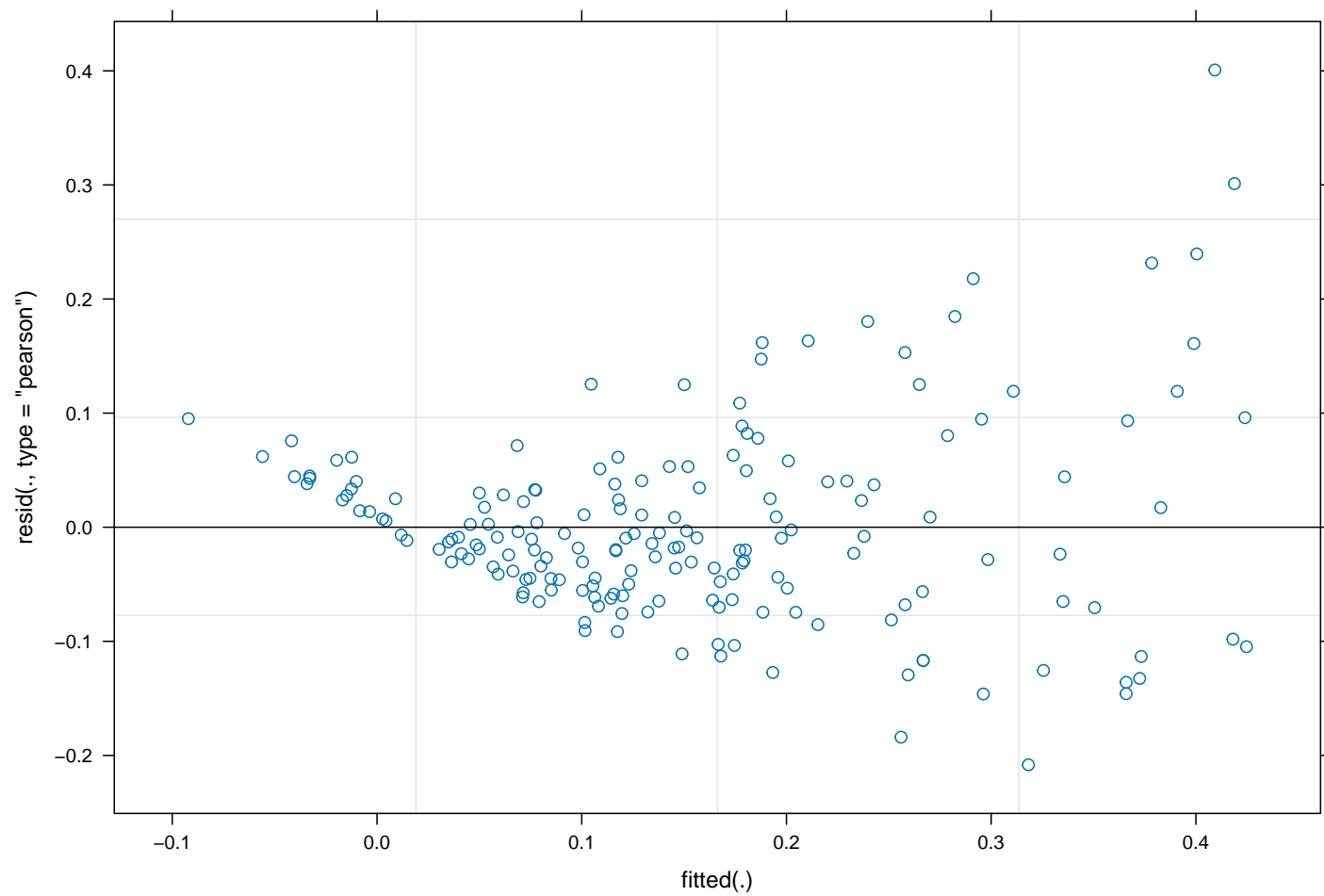
The estimated random slope standard deviations are small relative to that of the residual, which possibly incurred convergence failure of the model fit. Also, the variance of the fitted residual appears to increase for larger fitted values, which suggests some variance-stabilizing transformation of the variables are needed.

```
mod = lmer(accel ~ distance + soil + Richter + (1+ soil + distance|Quake),
           dat = dat, REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0752062 (tol = 0.002, component 1)
```

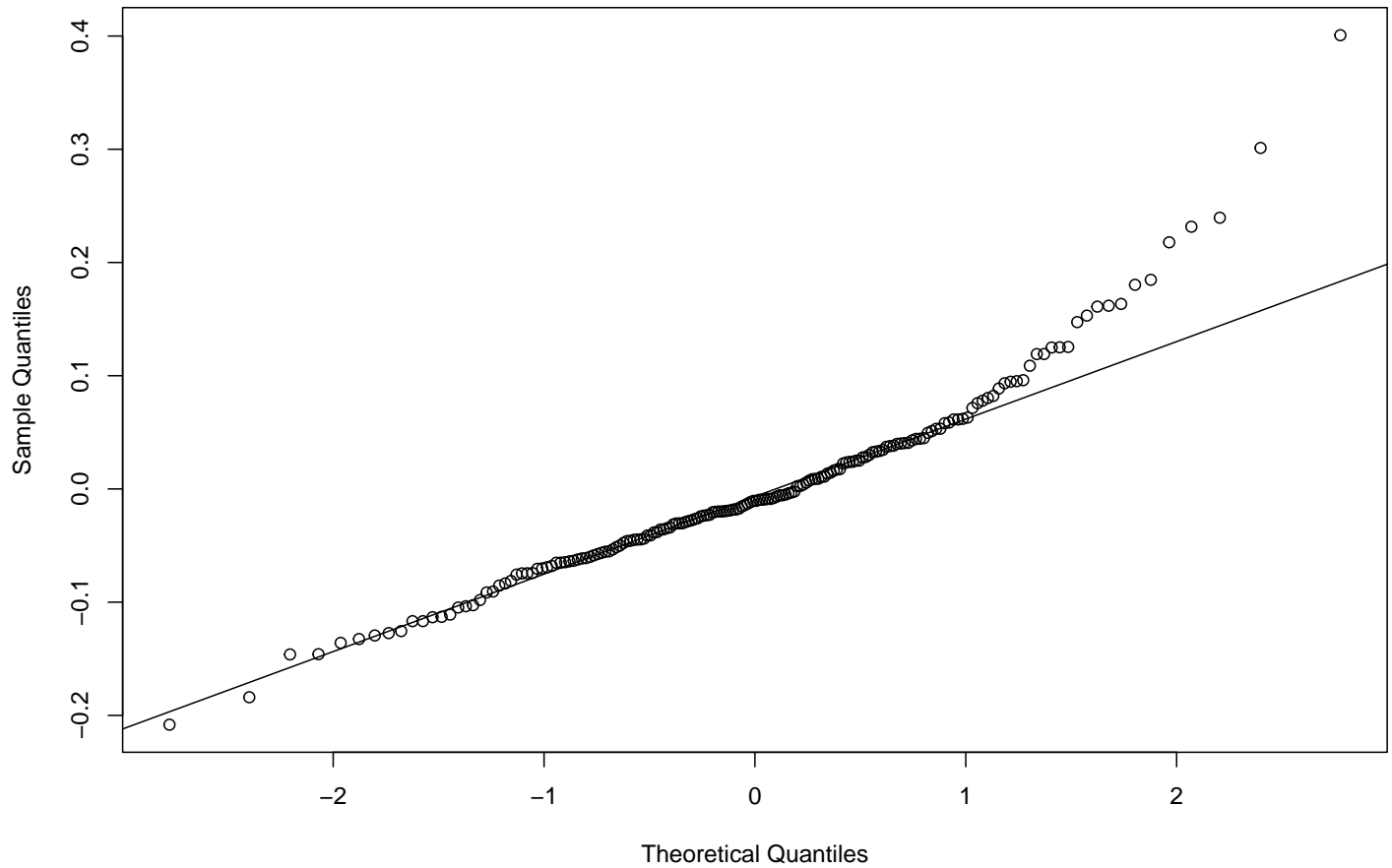
```
summary(mod)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: accel ~ distance + soil + Richter + (1 + soil + distance | Quake)
## Data: dat
##
##      AIC      BIC   logLik deviance df.resid
## -293.5   -258.3   157.8   -315.5     171
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.3101 -0.5872 -0.1167  0.4373  4.4464
##
## Random effects:
## Groups      Name                Variance Std.Dev. Corr
## Quake      (Intercept) 8.859e-03 0.094123
##           soil1      5.233e-05 0.007234 -0.46
##           distance  4.010e-06 0.002003 -0.91  0.27
## Residual              8.122e-03 0.090123
## Number of obs: 182, groups: Quake, 23
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.0510243  0.1142440  -0.447
## distance    -0.0030929  0.0006091  -5.078
## soil1        0.0084321  0.0207979   0.405
## Richter      0.0443811  0.0188413   2.356
##
## Correlation of Fixed Effects:
##              (Intr) distnc soil1
## distance    0.090
## soil1      -0.248 -0.039
## Richter    -0.965 -0.272  0.106
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0752062 (tol = 0.002, component 1)
plot(mod)
```



```
qqnorm(residuals(mod)); qqline(residuals(mod))
```

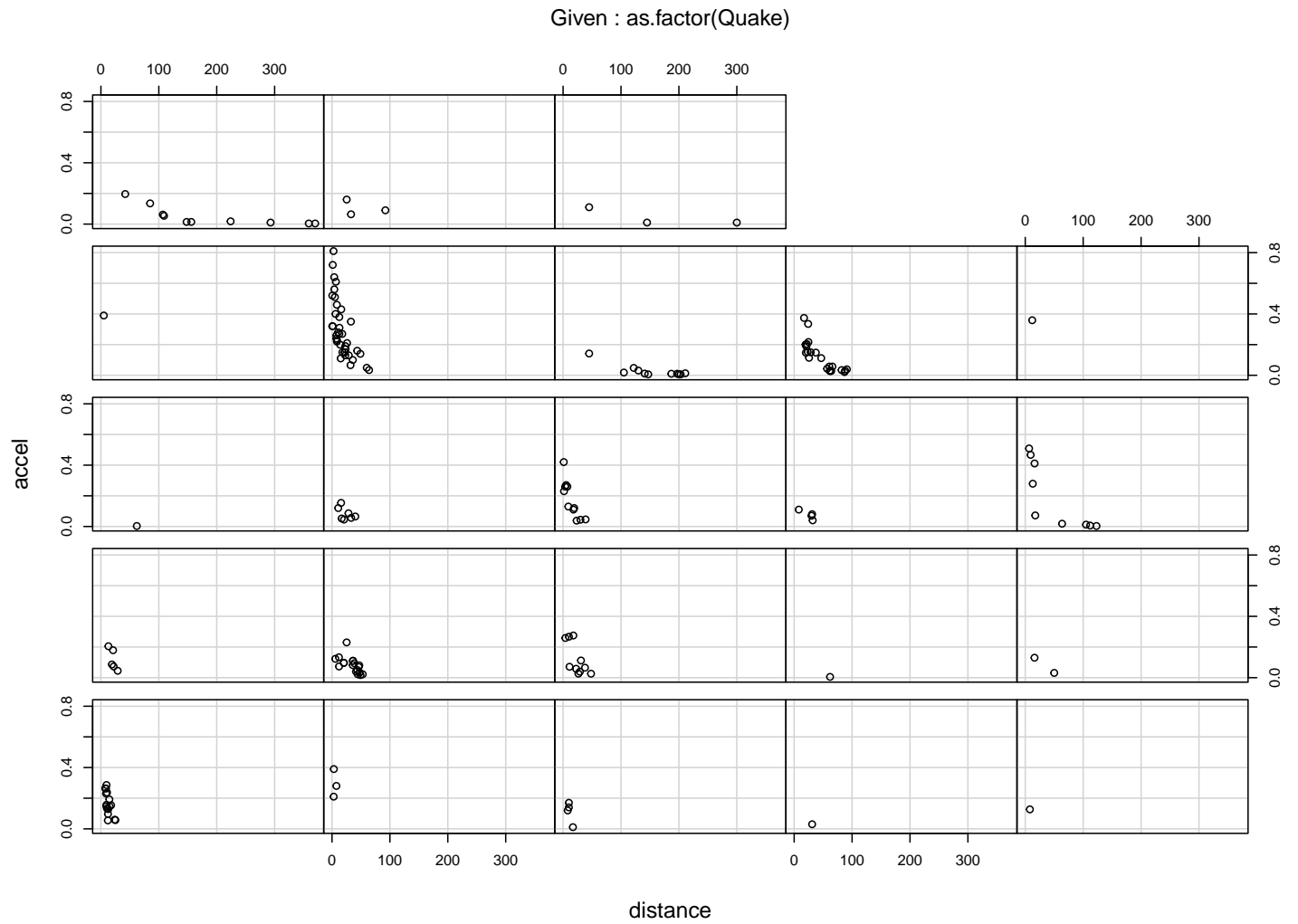
Normal Q–Q Plot



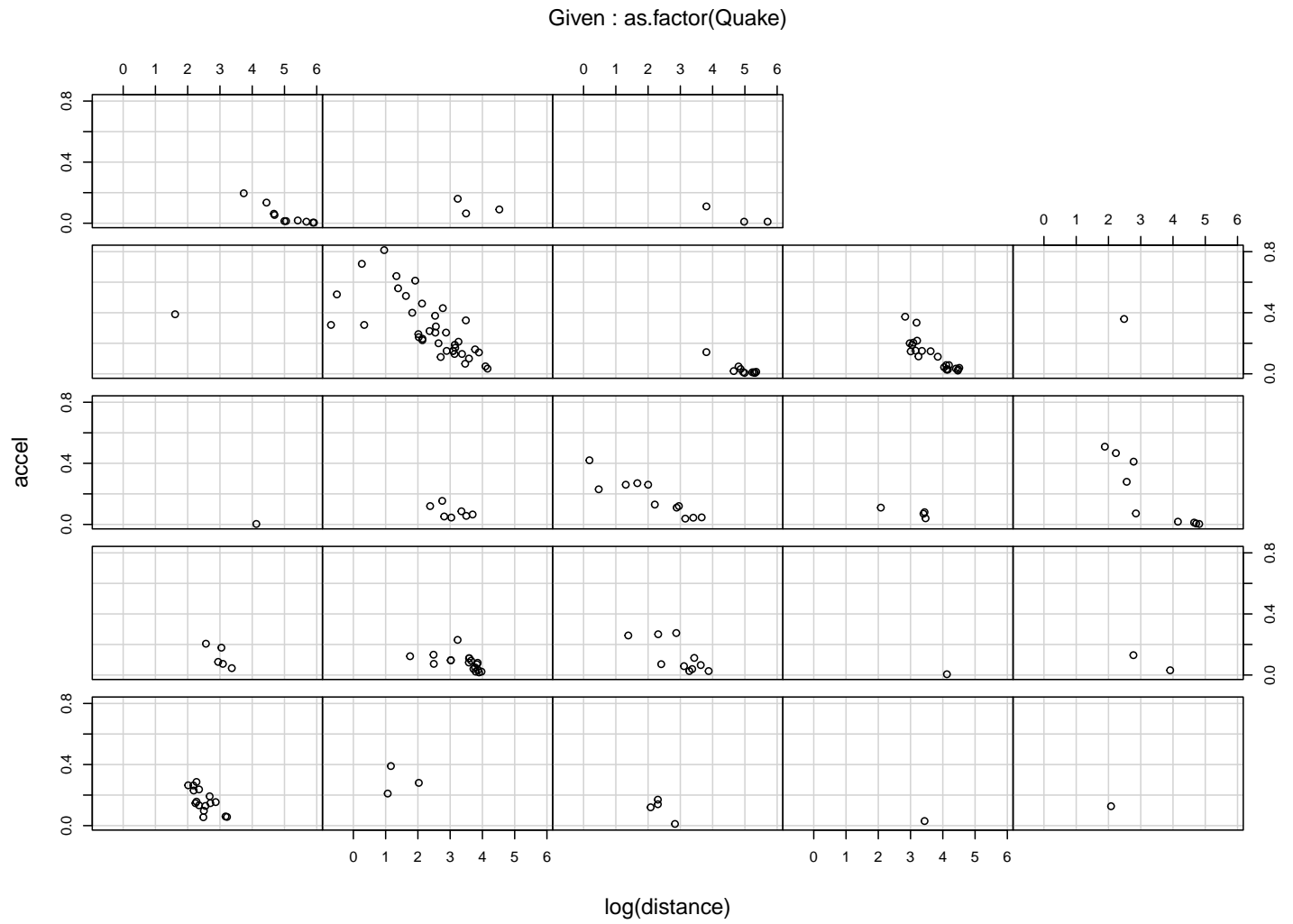
b.

Before transformation, the relationship does not look linear. After log transformation of accel and/or distance, it becomes closer to linear.

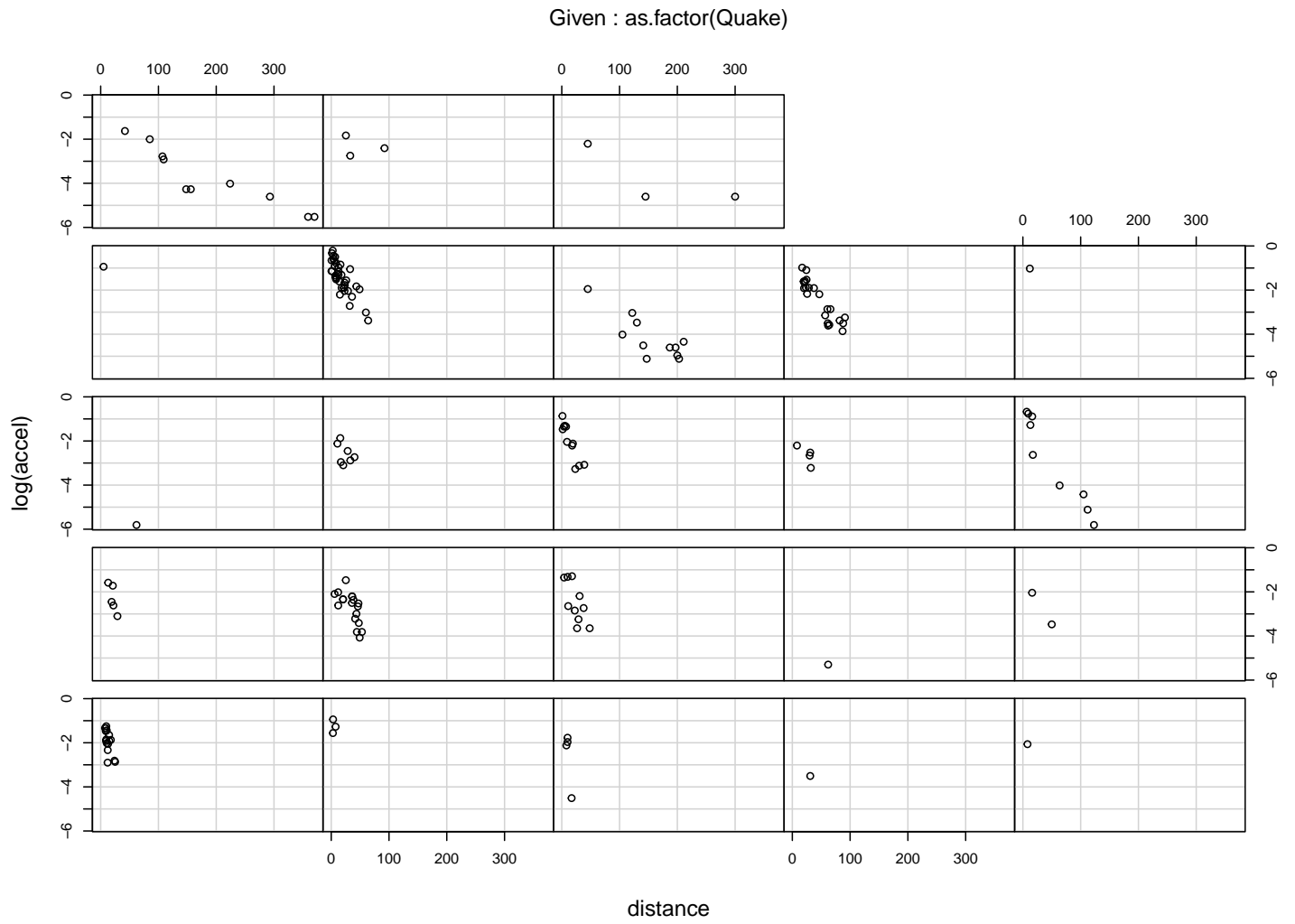
```
coplot(accel ~ distance | as.factor(Quake), data = dat, show.given=F)
```



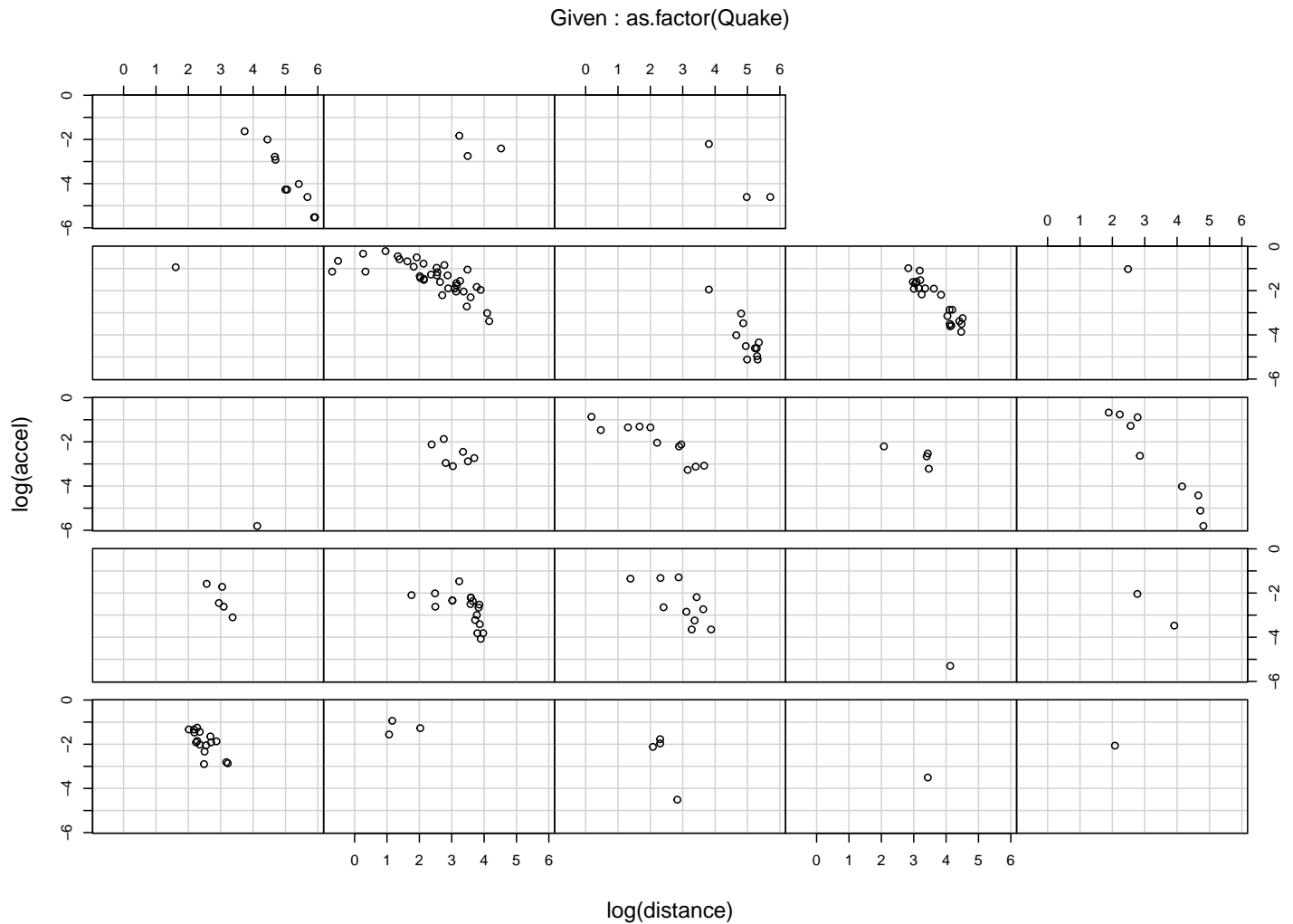
```
coplot(accel ~ log(distance) | as.factor(Quake), data = dat, show.given=F)
```



```
coplot(log(accel) ~ distance | as.factor(Quake), data = dat, show.given=F)
```



```
coplot(log(accel) ~ log(distance) | as.factor(Quake), data = dat, show.given=F)
```

c.

The residuals look closer to normality.

```
dat$taccel = log(dat$accel)
dat$tdist = log(dat$distance)
mod1 = lmer(taccel ~ tdist + soil + Richter + (1 + soil + tdist | Quake),
            dat = dat, REML=F)
```

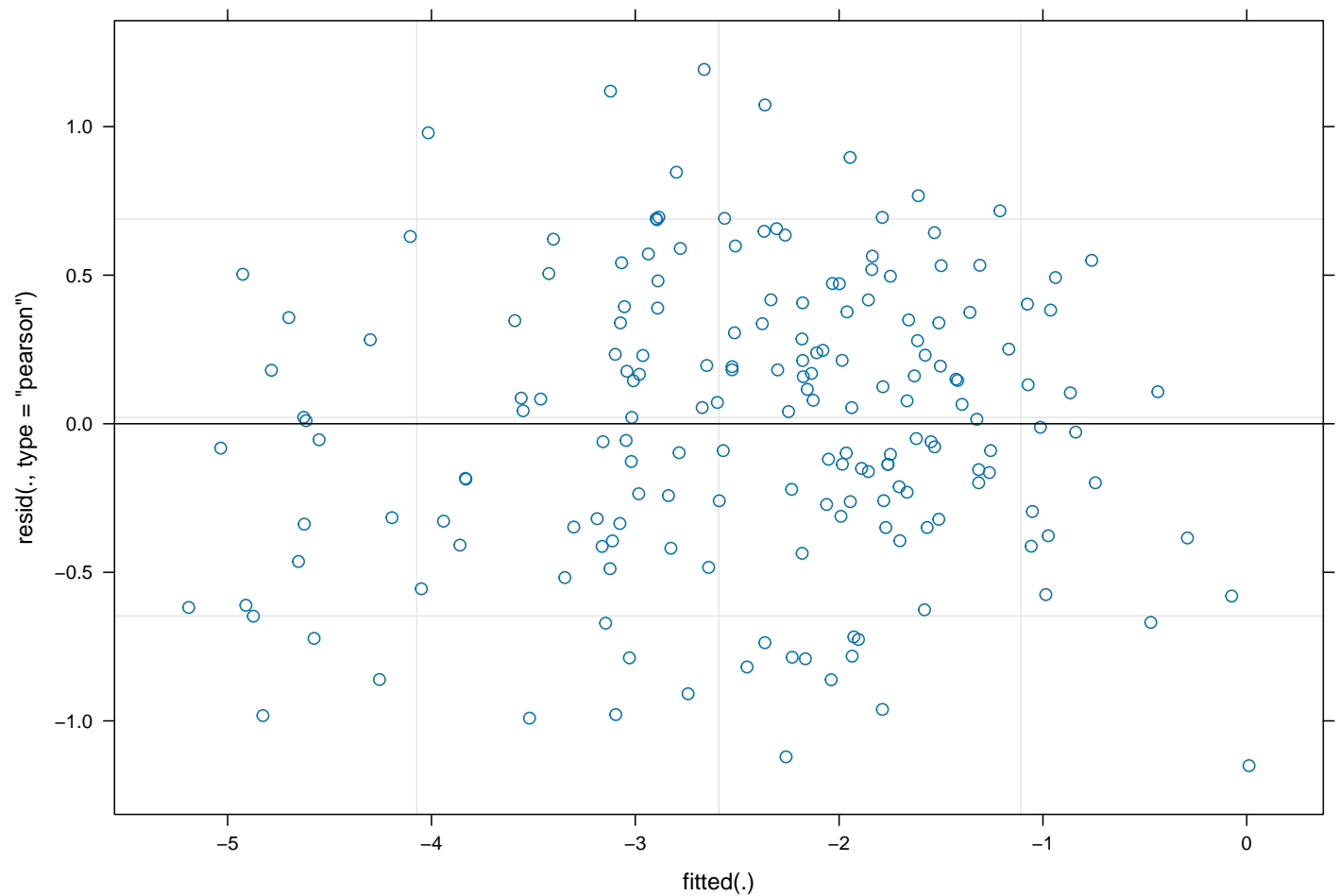
```
## boundary (singular) fit: see help('isSingular')
```

```
summary(mod1)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: taccel ~ tdist + soil + Richter + (1 + soil + tdist | Quake)
## Data: dat
##
##      AIC      BIC   logLik deviance df.resid
##  343.9   379.2  -161.0   321.9     171
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.24054 -0.63518  0.06088  0.67903  2.32193
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## Quake    (Intercept)  2.5550     1.5984
##          soil1        0.1877     0.4332  -0.78
```

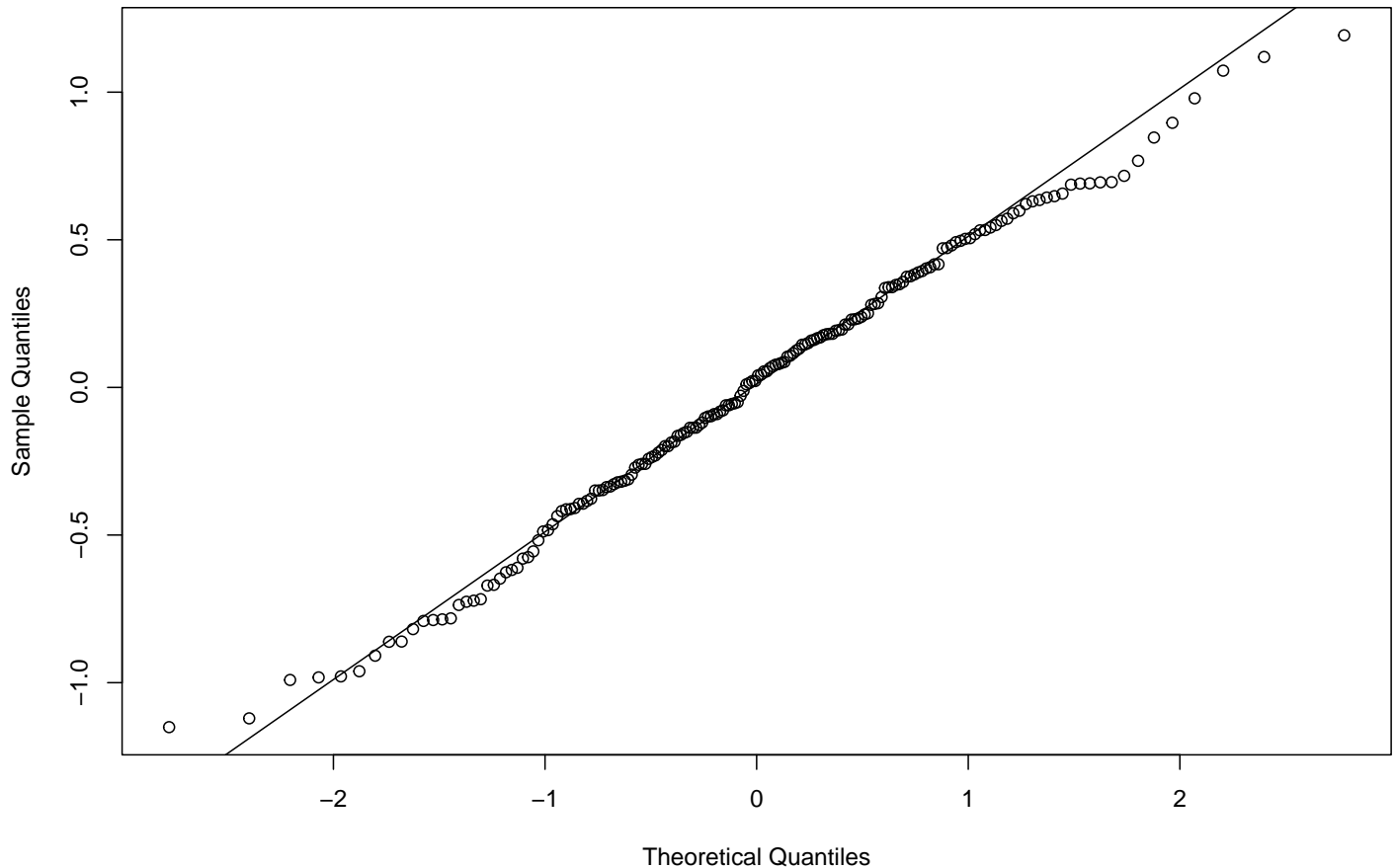
```
##          tdist      0.1984   0.4455   -0.96   0.57
## Residual      0.2638   0.5136
## Number of obs: 182, groups: Quake, 23
##
## Fixed effects:
##          Estimate Std. Error t value
## (Intercept) -2.54477   0.64779  -3.928
## tdist      -1.13725   0.11849  -9.598
## soil1       0.18809   0.17570   1.071
## Richter     0.59598   0.09415   6.330
##
## Correlation of Fixed Effects:
##      (Intr) tdist  soil1
## tdist  -0.457
## soil1  -0.376  0.283
## Richter -0.753 -0.177 -0.047
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
plot(mod1)
```



```
qqnorm(residuals(mod1)); qqline(residuals(mod1))
```

Normal Q-Q Plot



d.

We compare the full model in a. against two submodels, 1) without random slope tdist 2) without random slope soil. Since the full model has 3 random effects and the submodel has 2, the null distribution is a equal weight mixture of $\text{chisq}(df=3)$ and $\text{chisq}(df=2)$. According to LRT, the random slope tdist is significant but soil is not. BIC comparison leads to similar conclusion. Furthermore, starting from the full model in a., the LRTs for fixed effects by dropping one variable each time suggest that all except soil fixed effect are significant.

```
mod_full = lmer(taccel ~ tdist + soil + Richter + (1+ soil + tdist|Quake), dat = dat, REML=F)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
mod_nosoil = lmer(taccel ~ tdist + soil + Richter + (1+ tdist|Quake), dat = dat, REML=F)
```

```
mod_notdist = lmer(taccel ~ tdist + soil + Richter + (1+ soil|Quake), dat = dat, REML=F)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
lambda = c(2*(logLik(mod_full) - logLik(mod_nosoil)))
0.5*(1 - pchisq(lambda, 3)) + 0.5*(1 - pchisq(lambda, 2))
```

```
## [1] 0.08982154
```

```
lambda = c(2*(logLik(mod_full) - logLik(mod_notdist)))
0.5*(1 - pchisq(lambda, 3)) + 0.5*(1 - pchisq(lambda, 2))
```

```
## [1] 3.329181e-11
```

```
BIC(mod_full)
```

```
## [1] 379.1822
```

```

BIC(mod_nosoil) # BIC decreases (better model)

## [1] 369.3348
BIC(mod_notdist) # BIC increases (worse model)

## [1] 414.2665
mod_notdist = lmer(taccl ~ soil + Richter + (1 + soil + tdist|Quake),
  dat = dat, REML=F)

## boundary (singular) fit: see help('isSingular')
anova(mod_notdist, mod_full)

## Data: dat
## Models:
## mod_notdist: taccl ~ soil + Richter + (1 + soil + tdist | Quake)
## mod_full: taccl ~ tdist + soil + Richter + (1 + soil + tdist | Quake)
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## mod_notdist    10 376.77 408.81 -178.38   356.77
## mod_full       11 343.94 379.18 -160.97   321.94 34.831  1 3.596e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

mod_nosoil = lmer(taccl ~ tdist + Richter + (1 + soil + tdist|Quake),
  dat = dat, REML=F)

## boundary (singular) fit: see help('isSingular')
anova(mod_nosoil, mod_full)

## Data: dat
## Models:
## mod_nosoil: taccl ~ tdist + Richter + (1 + soil + tdist | Quake)
## mod_full: taccl ~ tdist + soil + Richter + (1 + soil + tdist | Quake)
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## mod_nosoil    10 342.89 374.93 -161.45   322.89
## mod_full       11 343.94 379.18 -160.97   321.94 0.9563  1 0.3281

mod_noRich = lmer(taccl ~ tdist + soil + (1 + soil + tdist|Quake),
  dat = dat, REML=F)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0040441 (tol = 0.002, component 1)

anova(mod_noRich, mod_full)

## Data: dat
## Models:
## mod_noRich: taccl ~ tdist + soil + (1 + soil + tdist | Quake)
## mod_full: taccl ~ tdist + soil + Richter + (1 + soil + tdist | Quake)
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## mod_noRich    10 364.26 396.30 -172.13   344.26
## mod_full       11 343.94 379.18 -160.97   321.94 22.326  1 2.301e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

e.

Summarize your answers leading upto a-d, and interpret the fixed effects coefficients estimates coherently. For example, for a unit change of tdist (multiplicative increase by e in the original scale), the log accel decreases by -1.13725.