Statistical Analysis of Designs with Repeated Measures by Linear Mixed Models

Livio Finos

Contents

1 Motivation

1.1 Import

```
load("./dataset/datiEEG_LMM_2x2.Rdata")
summary(dati)
```

```
##       Subj      Chan      Condition       Y
##   s01    : 16   O1:160   f:160     Min.   :-9.9964
##   s02    : 16   O2:160   n:160     1st Qu.:-2.5505
##   s03    : 16                      Median :-0.8299
##   s04    : 16                      Mean   :-0.6867
##   s05    : 16                      3rd Qu.: 1.3913
##   s06    : 16                      Max.   : 5.5468
##   (Other):224
```

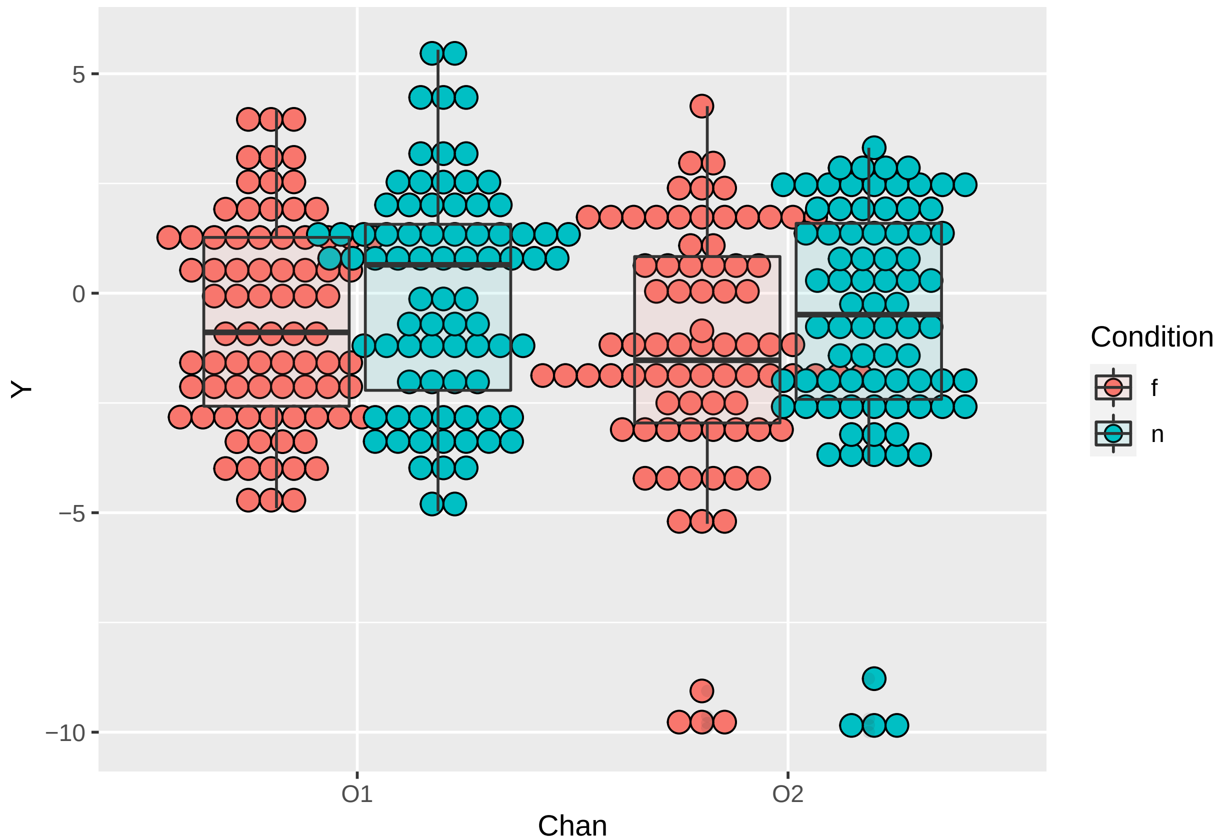
- 10 Subjects,
- 2-levels factor `Chan`
- 2-levels factor `Condition`

1.2 EDA

For Y: P300 Average Amplitude

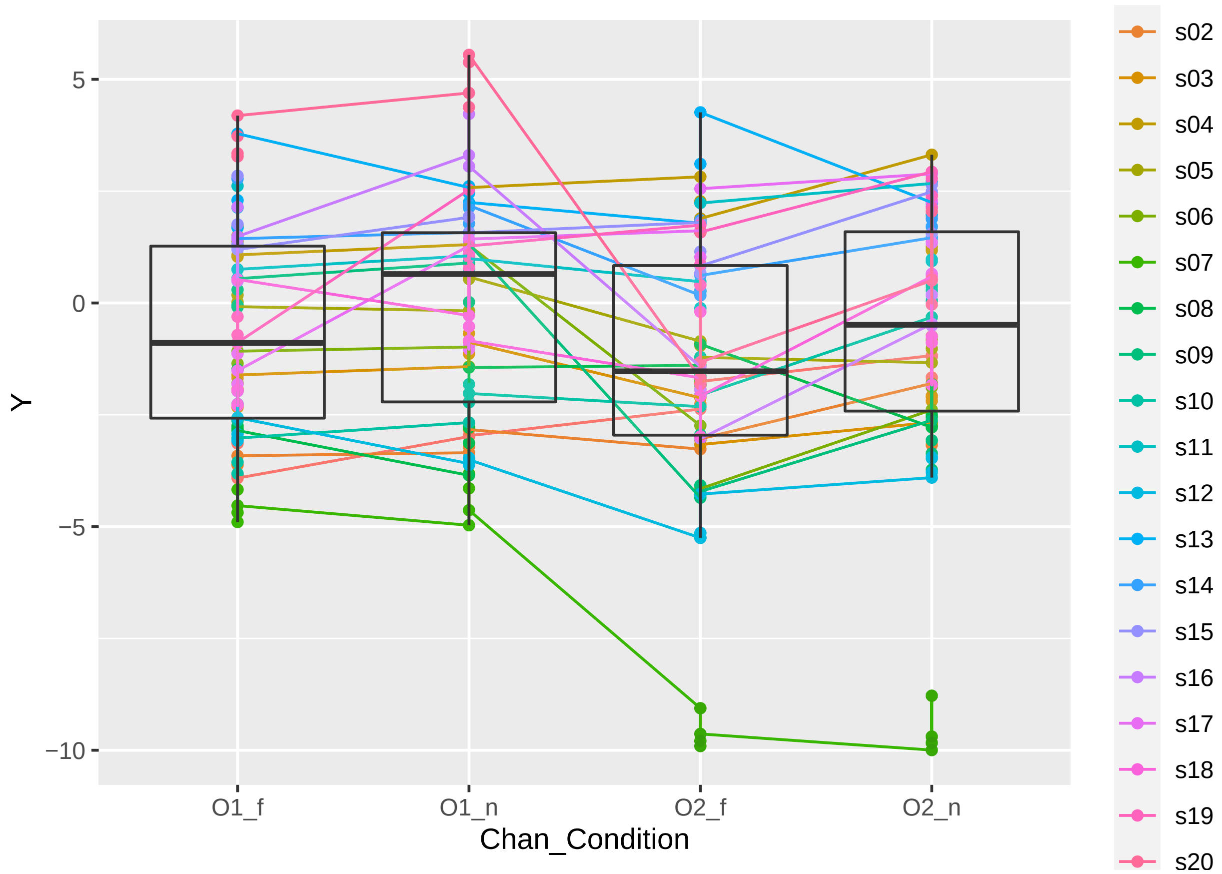
```
library(ggplot2)
p <- ggplot(dati,aes(x=Chan,y=Y,fill=Condition))
p=p+geom_dotplot(binaxis = "y",position=position_dodge(0.8),stackdir = "center") +geom_boxplot(alpha=.1)
p
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with `binwidth`.
```



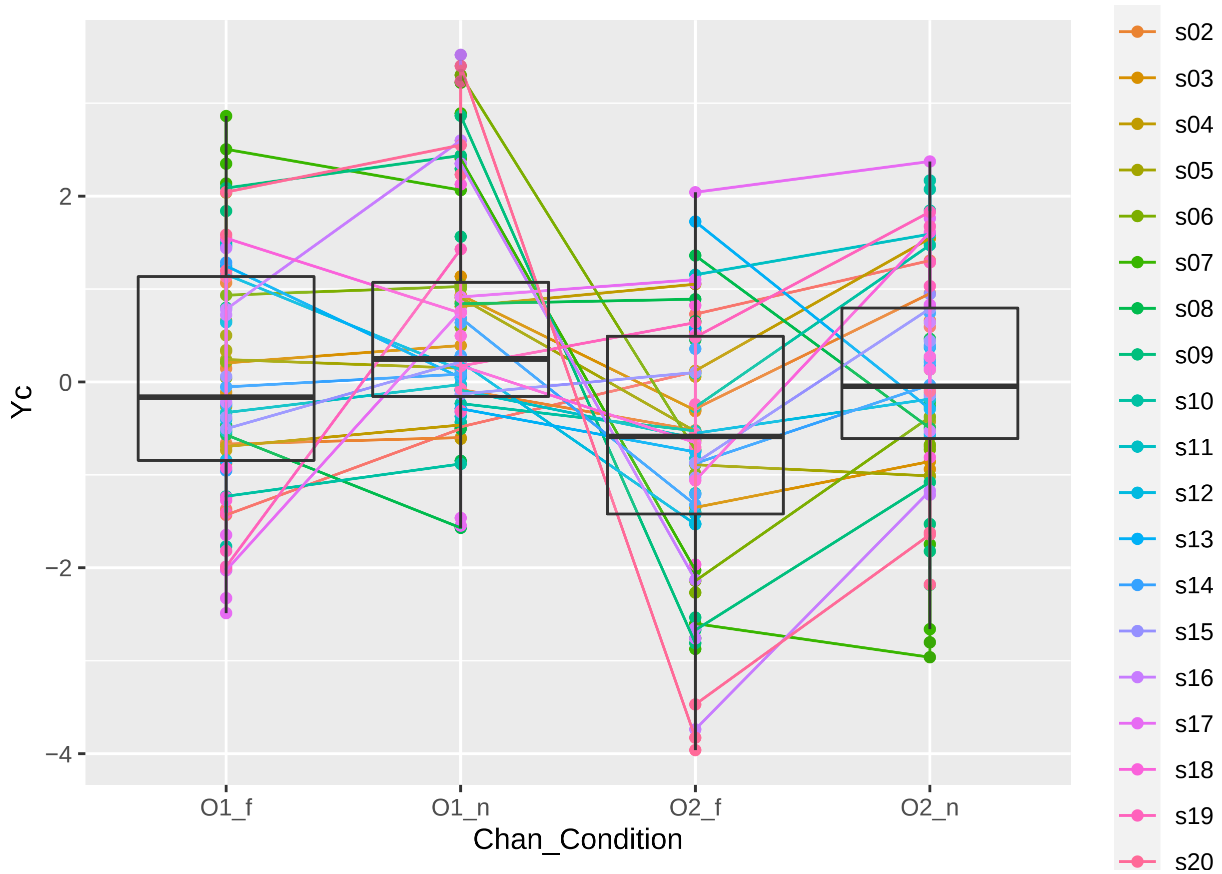
Is there a specificity of the Subject?

```
dati$Chan_Condition=paste(sep = "_",dati$Chan,dati$Condition)

p <- ggplot(dati,aes(x=Chan_Condition,y=Y))
p+geom_point(aes(group = Subj, colour = Subj))+
  geom_line(aes(group = Subj, colour = Subj))+
   geom_boxplot(alpha=.1)
```



We subtract the Subject-specific effect (i.e. Subject's mean) to each observation.

```
mod=lm(Y~Subj,data=dati)
# summary(mod)
Y=residuals(mod)
dati$Yc=as.vector(Y)



library(ggplot2)
p <- ggplot(dati,aes(Chan_Condition,Yc))
p+geom_point(aes(group = Subj, colour = Subj))+
  geom_line(aes(group = Subj, colour = Subj))+
   geom_boxplot(alpha=.1)
```



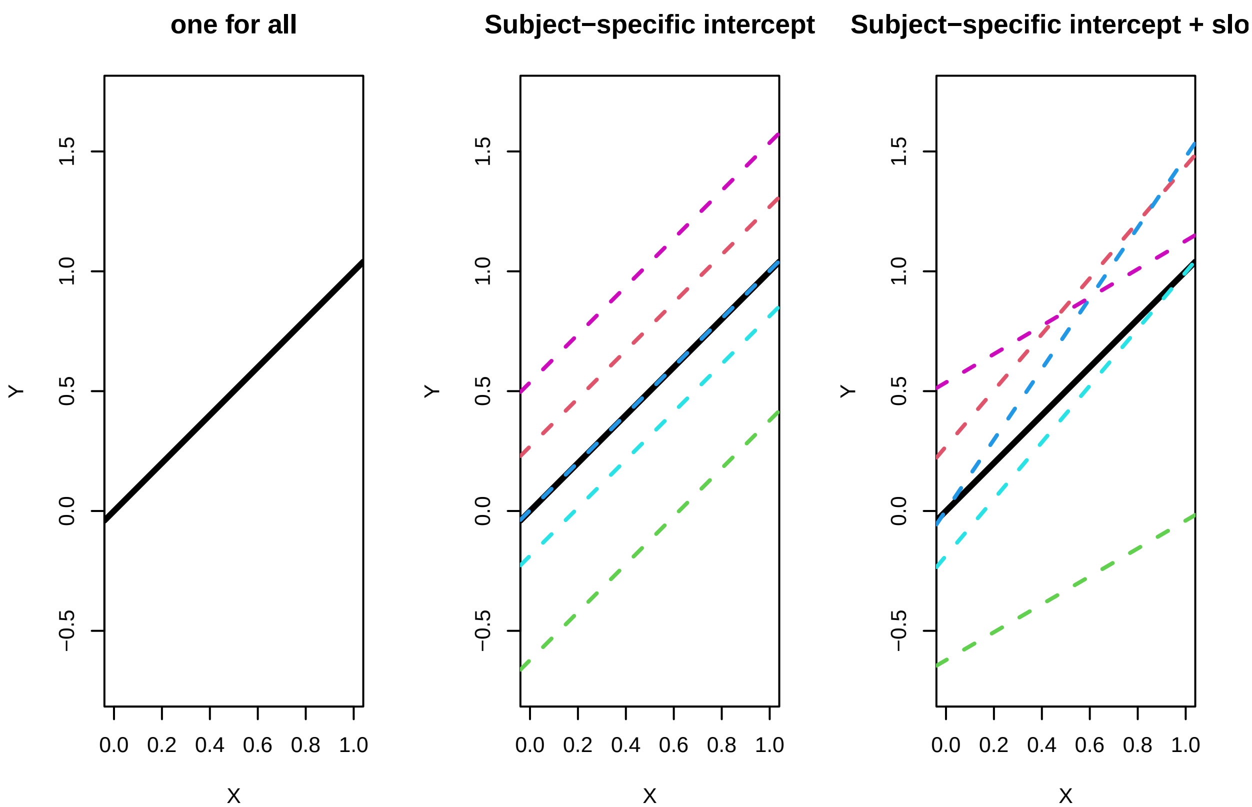
The dispersion of the data has been largely reduced. This effect is the one taken in account by the models for repeated measures.

2 Mixed models

2.1 IMPORTANT REMARK about contrasts in (mixed) linear models

```
# VERY IMPORTANT:
contrasts(dati$Condition) <- contr.sum(2) #2 is the number of levels
contrasts(dati$Chan) <- contr.sum(2) #2 is the number of levels
```

2.2 Intuition



Mixed models allow to model Subject-specific (average) effect by assuming that it is randomly drawn from the dstribution of the population (which is normal).

I assume you are expert on mixed models, if not https://en.wikipedia.org/wiki/Mixed_model
and much more on: http://webcom.upmf-grenoble.fr/LIP/Perso/DMuller/M2R/R_et_Mixed/documents/Bates-book.pdf
and
https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf

Due to the small size of the dataset, in our example we only explore the scenario with random intercept and fixed slope (i.e. a simpler model, less parameters).

2.3 The model

Models with random effects can be defined as:

$$Y_{n\times 1} = X_{n\times p}B_{p\times 1} + Z_{n\times q}b_{q\times 1} + \varepsilon_{n\times 1}$$

where

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

In the models we will consider, the random effects are modeled as a multivariate normal random variable:

$$b \sim \mathcal{N}(0, \Sigma_{q\times q}),$$

In a *linear mixed model* the Conditional distribution $(Y|\mathcal{B} = b)$ is a *spherical* multivariate Gaussian.

In our case $n = \#Subj \times \#Chan \times \#Condition = 10 \times 2 \times 2 = 40$. X is the matrix of (dummified) predictors. Z can take many dimensions and values. Examples follow.

2.4 Different Models

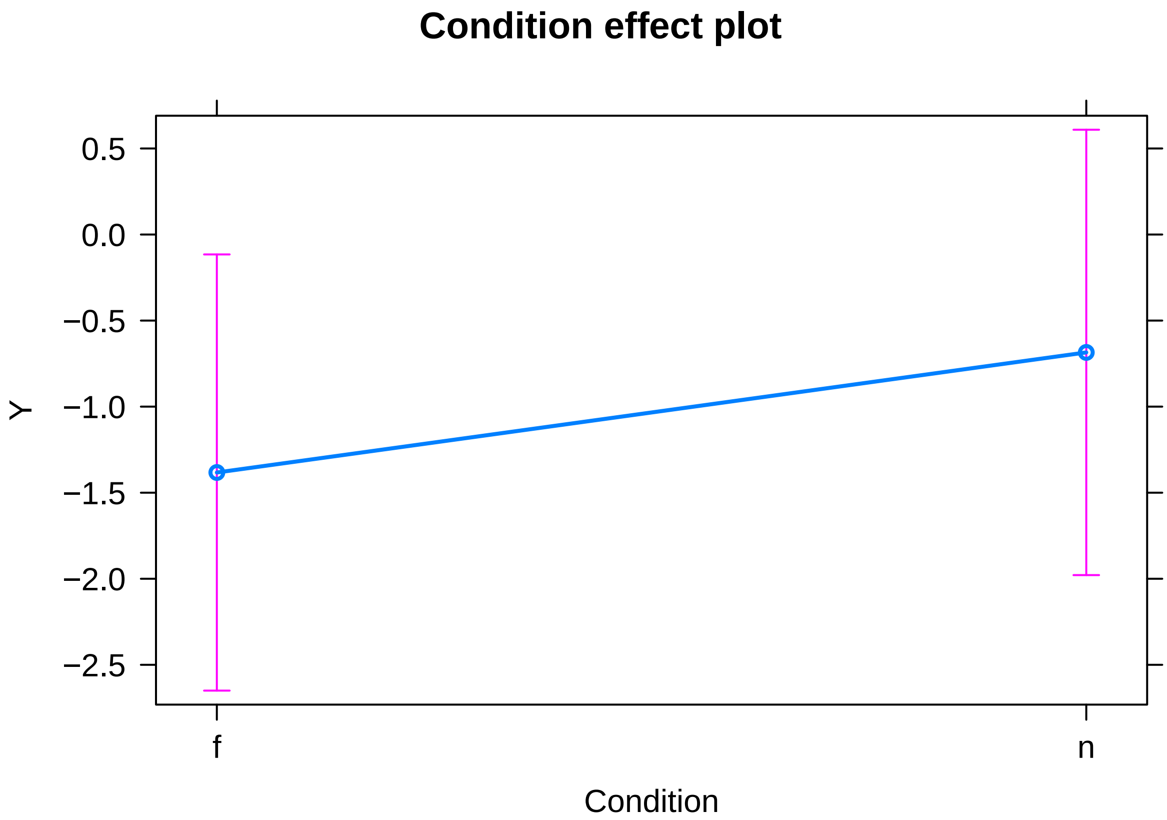
2.4.1 ONLY Condition, restricted to Chan=O2 (i.e. 1 Factor)

```
library(lmerTest)

mod=lmer(Y~ Condition +(1+Condition|Subj),data=subset(dati,Chan=="O2"))

summary(mod)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Y ~ Condition + (1 + Condition | Subj)
##    Data: subset(dati, Chan == "O2")
##
## REML criterion at convergence: 456.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8843 -0.6100 -0.0839  0.5110  3.2277
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subj     (Intercept) 8.1738   2.8590
##           Condition1  0.1115   0.3339   -0.09
##  Residual             0.4838   0.6955
## Number of obs: 160, groups:  Subj, 20
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept) -1.03388    0.64165 19.00049  -1.611  0.12361
## Condition1  -0.34883    0.09273 18.99999  -3.762  0.00132 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr)
## Condition1 -0.073
```

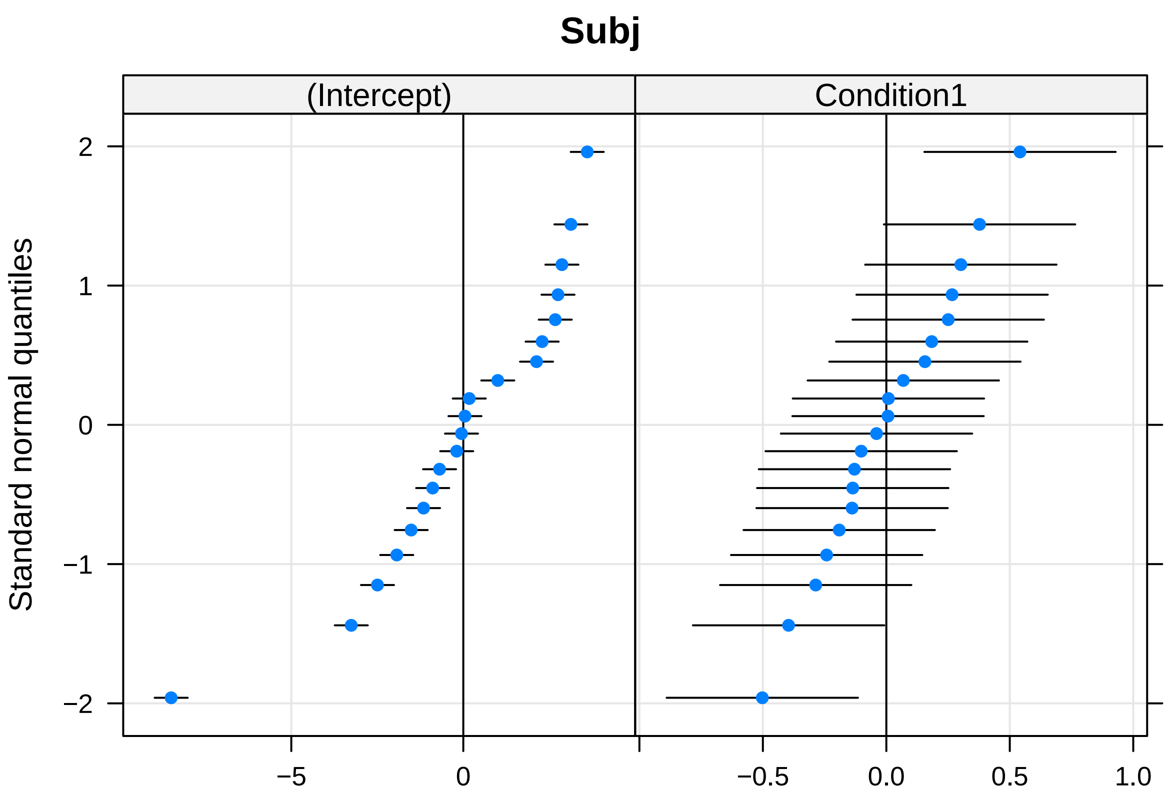
Plotting tools:

```
library(effects)
plot(allEffects(mod))
```




```
#plot random effects:
require(lattice)
qqmath(ranef(mod, condVar=TRUE))
```

```
## $Subj
```



2.4.2 Condition and Chan (i.e. 2 Factors)

Random effect for Subject (Random Intercept)

Z is the matrix of dummy variables of the column `dati$Subj`.

```
library(lmerTest)

mod2=lmer(Y~ Condition*Chan +(1|Subj),data=dati)

summary(mod2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Y ~ Condition * Chan + (1 | Subj)
##    Data: dati
##
## REML criterion at convergence: 1165
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.45740 -0.57314 -0.04347  0.69734  2.21259
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
```

```
##  Subj     (Intercept) 5.647    2.376
##  Residual             1.709    1.307
## Number of obs: 320, groups:  Subj, 20
##
## Fixed effects:
##                     Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)         -0.68675    0.53639  19.00000  -1.280    0.216
## Condition1          -0.32723    0.07309 297.00000  -4.477 1.08e-05 ***
## Chan1                0.34713    0.07309 297.00000   4.749 3.18e-06 ***
## Condition1:Chan1     0.02160    0.07309 297.00000   0.295    0.768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Cndtn1 Chan1
## Condition1  0.000
## Chan1       0.000  0.000
## Cndtn1:Chn1 0.000  0.000  0.000
```

```
car::Anova(mod2,type=3,test="F")
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: Y
##                       F Df Df.res    Pr(>F)
## (Intercept)      1.6392  1     19    0.2158
## Condition       20.0447  1    297 1.08e-05 ***
## Chan            22.5567  1    297 3.18e-06 ***
## Condition:Chan   0.0873  1    297    0.7678
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Random effect for Subject: Random Intercept + radom factors

```
library(lmerTest)

# A trick to make the model fit working. But take care of the (slight) different interpretation: coeffi

contrasts(dati$Condition)=contrasts(dati$Condition)/sqrt(nrow(dati))
contrasts(dati$Chan)=contrasts(dati$Chan)/sqrt(nrow(dati))

mod3=lmer(Y~ Condition*Chan +(1+Condition+Chan|Subj),data=dati)

# It could also work with +(1+Condition+Chan|Subj)
# but we simplify here the problem.

summary(mod3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Y ~ Condition * Chan + (1 + Condition + Chan | Subj)
##    Data: dati
```

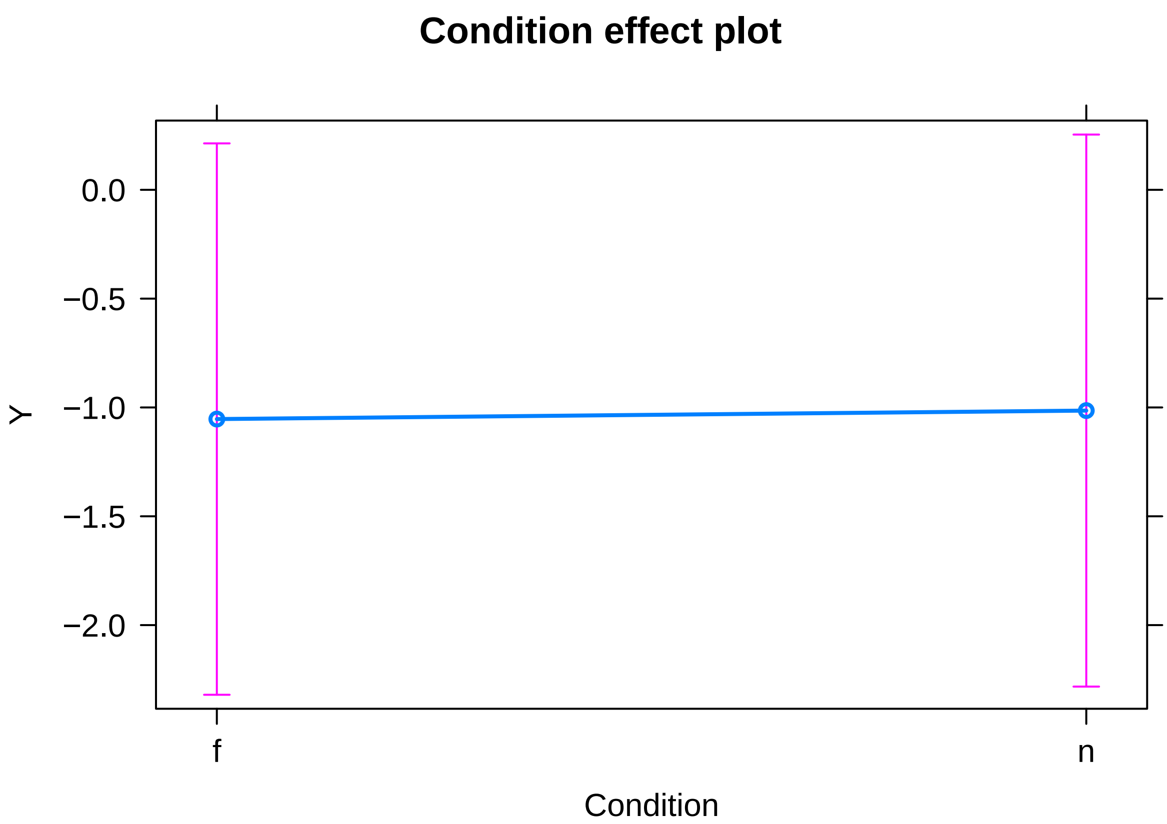
```
##
## REML criterion at convergence: 857.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1500 -0.6171 -0.1416  0.4583  3.2781
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subj     (Intercept)   5.7243  2.3926
##           Condition1   29.7257  5.4521  -0.30
##           Chan1       354.3697 18.8247  -0.27 -0.11
##  Residual               0.4808  0.6934
## Number of obs: 320, groups:  Subj, 20
##
## Fixed effects:
##                   Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)        -0.6867     0.5364  18.9992  -1.280 0.215849
## Condition1         -5.8536     1.4025  18.9995  -4.174 0.000515 ***
## Chan1               6.2096     4.2661  19.0013   1.456 0.161832
## Condition1:Chan1    6.9110    12.4038 259.0001   0.557 0.577893
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Cndtn1 Chan1
## Condition1  -0.258
## Chan1       -0.262 -0.091
## Cndtn1:Chn1  0.000  0.000  0.000
```

```
car::Anova(mod3,type=3,test="F")
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: Y
##                      F Df Df.res    Pr(>F)
## (Intercept)     1.6392  1     19 0.2158482
## Condition      17.4193  1     19 0.0005154 ***
## Chan            2.1187  1     19 0.1618332
## Condition:Chan  0.3104  1    259 0.5778929
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

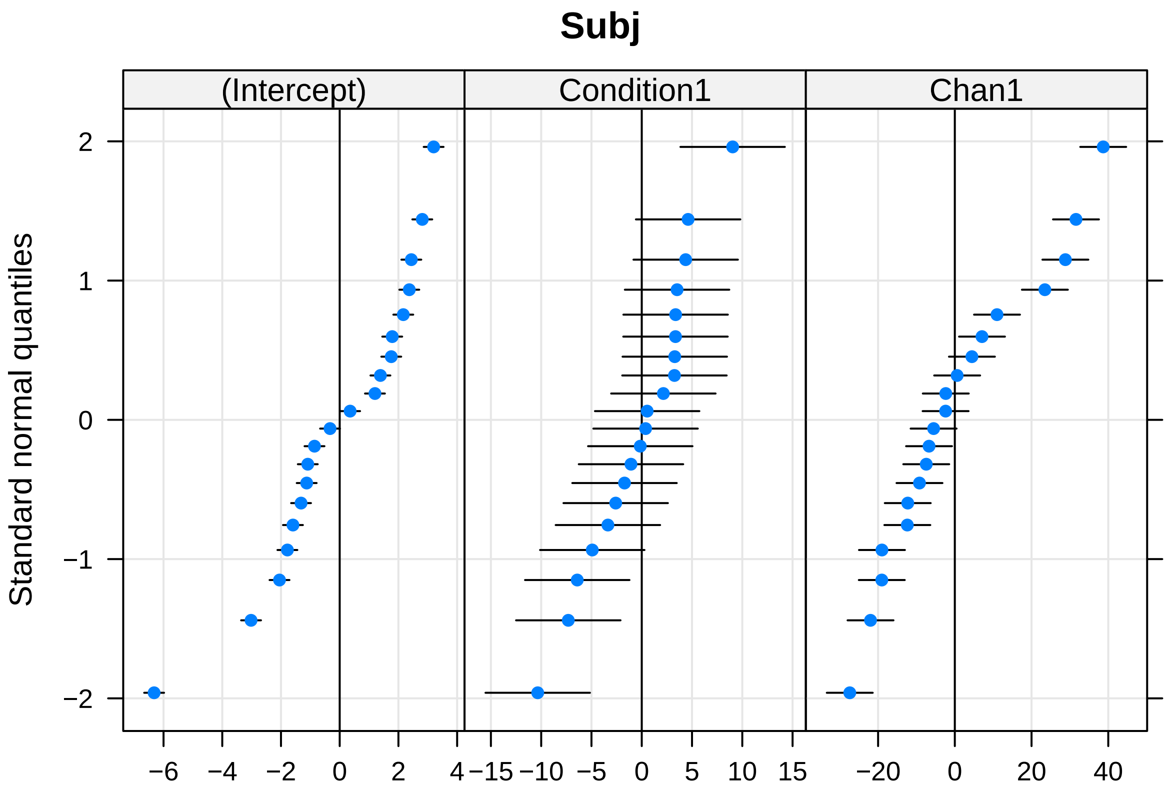
2.5 Plotting tools

```
library(effects)
plot(allEffects(mod))
```



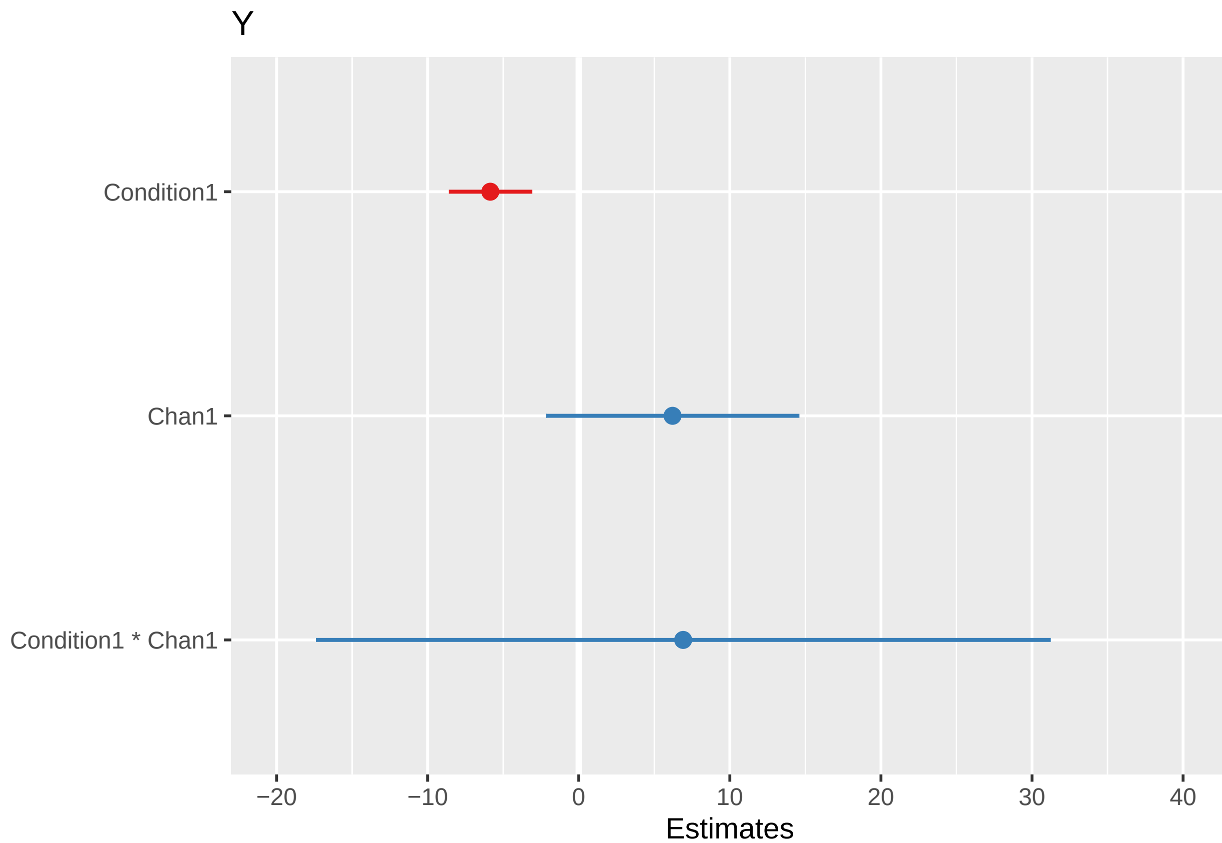

```r
#plot random effects:
require(lattice)
qqmath(ranef(mod3, condVar=TRUE))
```

```
## $Subj
```

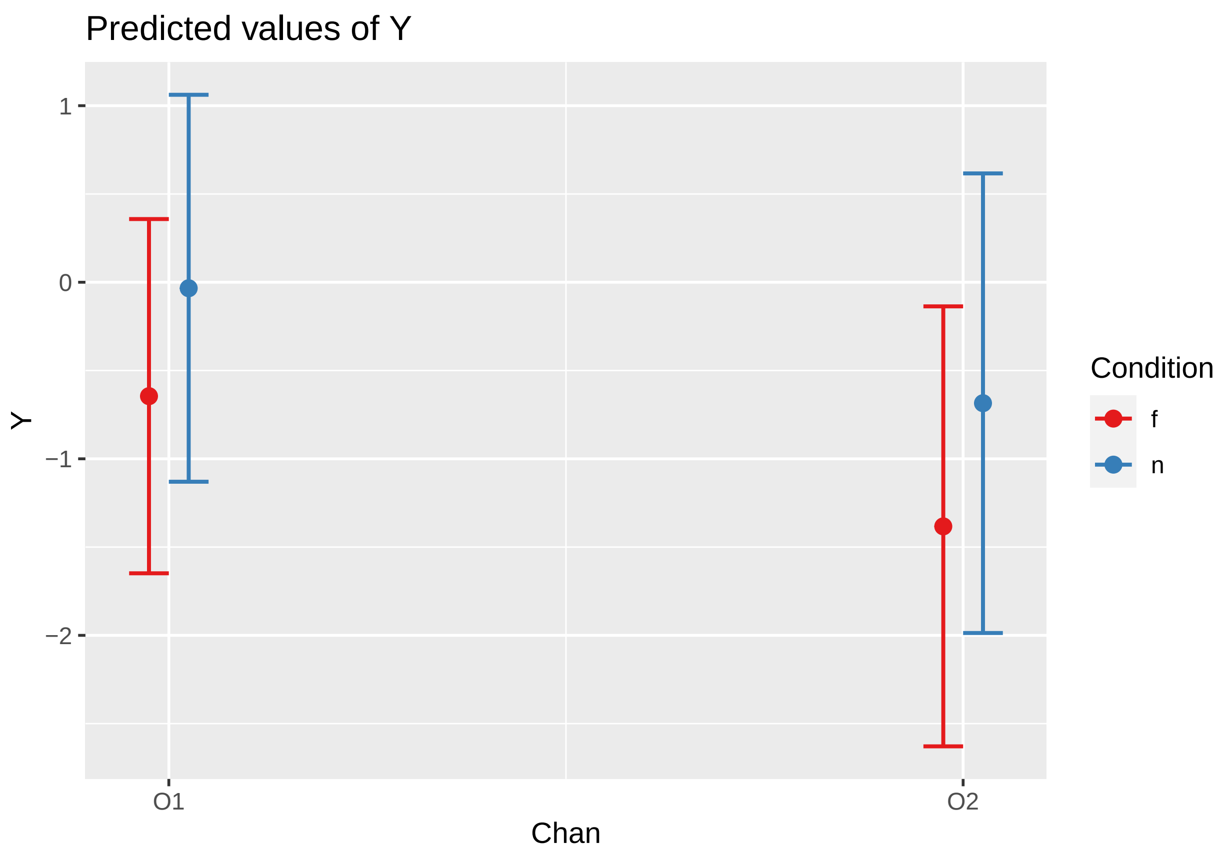


An alternative plotting tool:

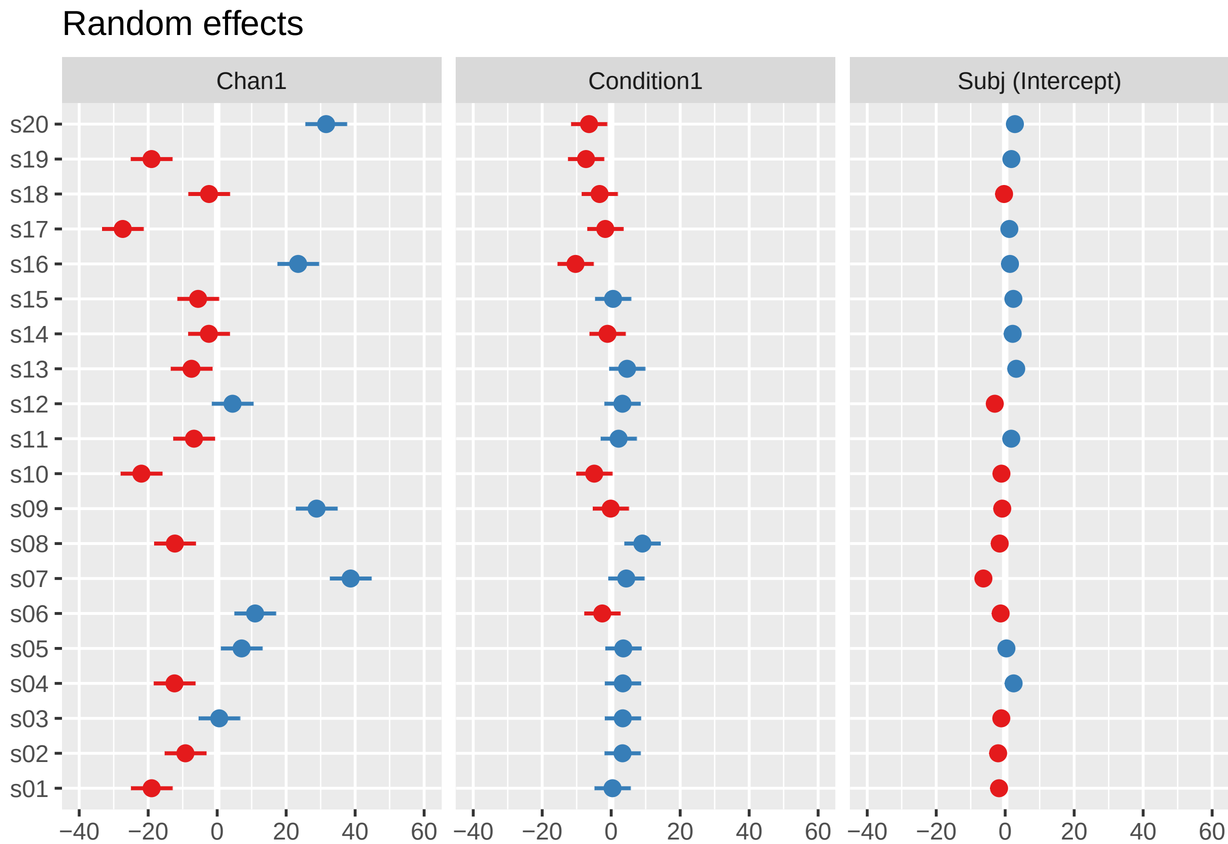
```
library(sjPlot)
library(ggplot2)
plot_model(mod3, type = "est")
```




```
plot_model(mod3, type = "eff", terms = c("Chan","Condition"))
```




```
plot_model(mod3, type = "re", terms = c("Chan","Condition"))
```



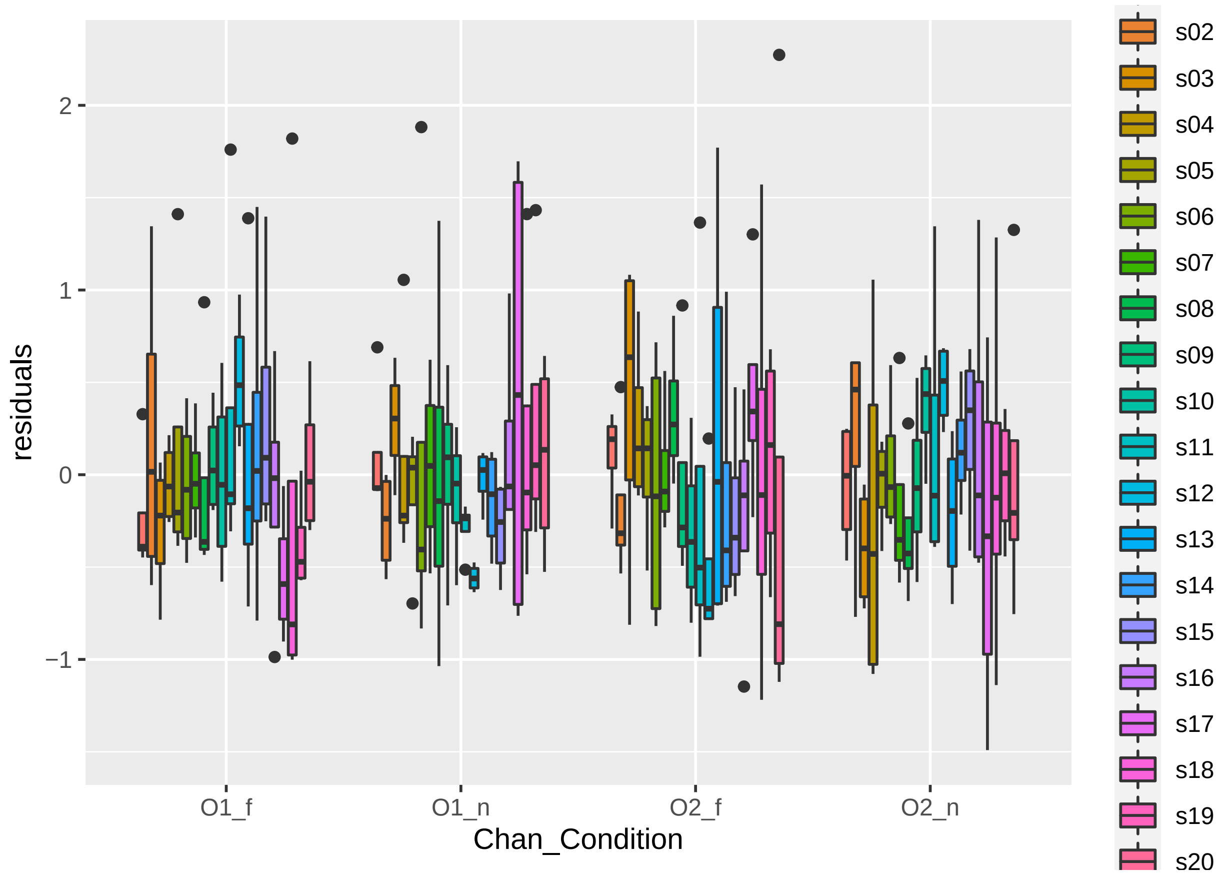
2.6 Validity of the assumptions

- Independence of the residuals?
- Normality of the residuals?
- Homoscedasticity of the residuals (i.e. same variance between Subject/Condition/Chan?
- outliers?
- Leaverage? (influential observations)

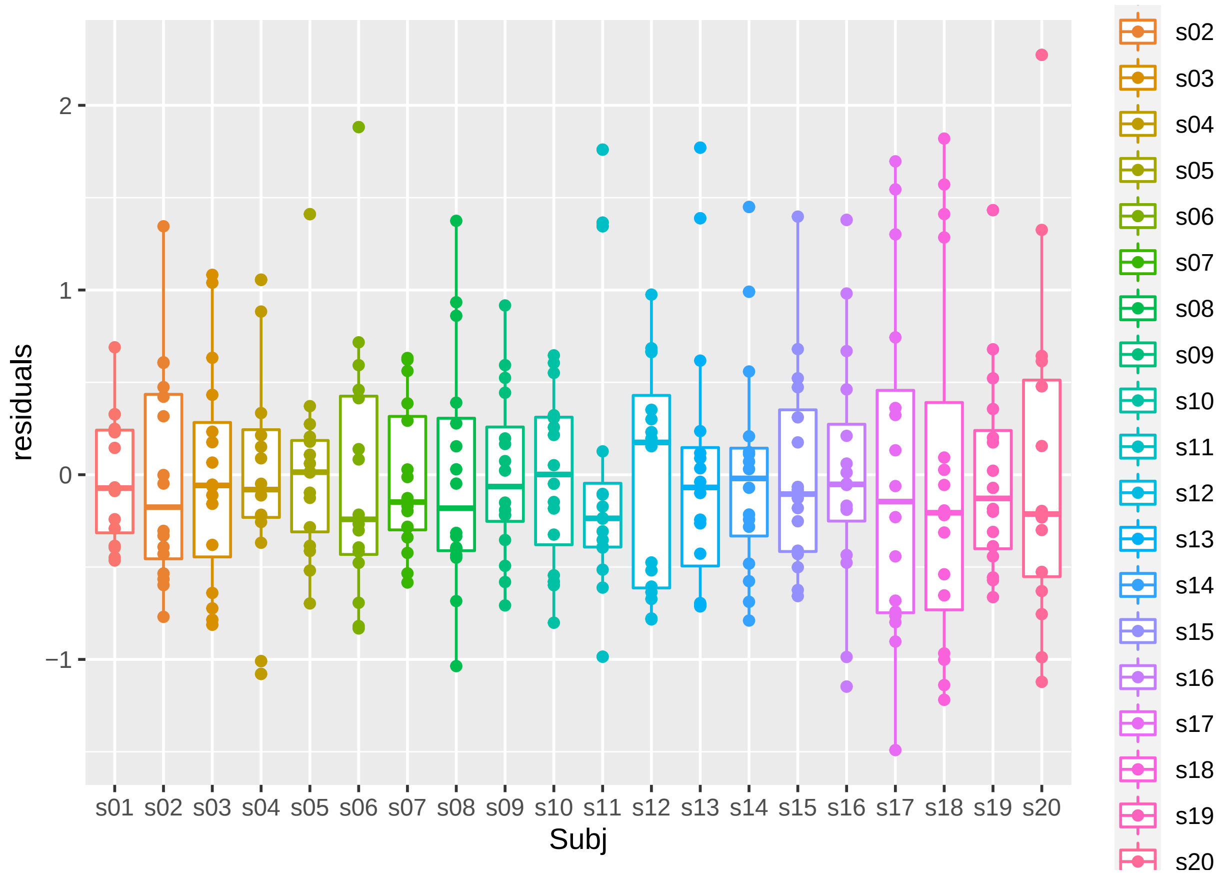
Please, do not test for normality, for homoscedasticity, sphericity etc.

Use Exploratory data Analysis, instead!

```
dati$residuals=residuals(mod3)
p <- ggplot(dati, aes(x=Chan_Condition, y=residuals,fill=Subj)) +   geom_boxplot()
p
```




```
p <- ggplot(dati, aes(x=Subj, y=residuals,col=Subj)) +   geom_boxplot()+  geom_point(aes(group = interac
p
```



3 (minimal) Bibliography

Jonathan Baron (2011) Notes on the use of R for psychology experiments and questionnaires https://www.sas.upenn.edu/~baron/from_cattell/rpsych/rpsych.html

and Course materal of
ST 732, Applied Longitudinal Data Analysis, NC State University by Marie Davidian https://www.stat.ncsu.edu/people/davidian/courses/st732/notes/chap5.pdf from https://www.stat.ncsu.edu/people/davidian/courses/st732/

About Type I, II, III SS: https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained/

About Mixed models:
http://webcom.upmf-grenoble.fr/LIP/Perso/DMuller/M2R/R_et_Mixed/documents/Bates-book.pdf
and
https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf