Orientation: L2

## Data analysis

We load the data and remove the cases with NA values.

We have 122 subjects for a total of 1952 observations.

After removal of entries with missing data due to wrong answers, we have 118 subjects and 1476 observations left.

We consider the following independent variables: - LEAYRS a numerical variable with values ranging from 1 to 16. Learning in years. - POST1 a percentage scaled to lay between 0 and 10. It indicates the post test proficency. - PRE a percentage scaled to lay between 0 and 10. It indicates the previous knowledge. - AMSP a numerical variable with values ranging from 1 to 5. - HRSD a numerical variable with values ranging from 0 to 11.5. L2 use in hours per day.

We rescaled some of these variables to be on similar scales. We treat these as fixed factors under study. In addition we have the random factors described earlier. The same modelling set up applies here.

Our base model is the following lmm <- lmer(log(Time) ~ type + LEAYRS + (PRE + POST1)^2 + AMSP + HRSD + (1 | List:Name) + (1 | List:type:Order), data = DLM)

lmm <- lmer(log(Time) ~ type + LEAYRS + (PRE + POST1)^2 + AMSP + HRSD + (1 | List:Name) + (1 | List:type:Order), data = DLM, REML=FALSE)  
summary(lmm)

## Linear mixed model fit by maximum likelihood ['lmerMod']  
## Formula:   
## log(Time) ~ type + LEAYRS + (PRE + POST1)^2 + AMSP + HRSD + (1 |   
## List:Name) + (1 | List:type:Order)  
## Data: DLM  
##   
## AIC BIC logLik deviance df.resid   
## 1866.8 1925.1 -922.4 1844.8 1465   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.134 -0.664 -0.098 0.580 3.087   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## List:Name (Intercept) 0.0533 0.231   
## List:type:Order (Intercept) 0.0206 0.143   
## Residual 0.1736 0.417   
## Number of obs: 1476, groups: List:Name, 118; List:type:Order, 32  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 7.81428 0.35869 21.79  
## typeMISMATCH -0.01946 0.05540 -0.35  
## LEAYRS -0.04039 0.01240 -3.26  
## PRE -0.18136 0.07570 -2.40  
## POST1 -0.11620 0.03554 -3.27  
## AMSP 0.04397 0.03757 1.17  
## HRSD 0.01637 0.01315 1.24  
## PRE:POST1 0.02002 0.00786 2.55  
##   
## Correlation of Fixed Effects:  
## (Intr) tMISMA LEAYRS PRE POST1 AMSP HRSD   
## typMISMATCH -0.076   
## LEAYRS -0.011 0.000   
## PRE -0.852 -0.004 -0.005   
## POST1 -0.887 -0.001 -0.067 0.853   
## AMSP -0.393 0.000 -0.101 0.068 0.017   
## HRSD 0.120 -0.001 -0.219 -0.006 0.023 -0.338   
## PRE:POST1 0.855 0.005 0.022 -0.986 -0.893 -0.069 -0.027

We use a manual, step-forward procedure with likelihood ratio test to see which of the fixed effects are significant.

lmm.0 <- lmer(log(Time) ~ (1 | List:Name) + (1 | List:type:Order), data = DLM)  
  
lmm.1 <- update(lmm.0, .~. + type)  
anova(lmm.0,lmm.1)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.1: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + type  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)  
## lmm.0 4 1876 1897 -934 1868   
## lmm.1 5 1878 1904 -934 1868 0.11 1 0.74

lmm.2 <- update(lmm.0, .~.+LEAYRS)  
anova(lmm.0,lmm.2)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.2: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + LEAYRS  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## lmm.0 4 1876 1897 -934 1868   
## lmm.2 5 1870 1896 -930 1860 8.04 1 0.0046 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

lmm.3 <- update(lmm.0, .~. + POST1)   
anova(lmm.0,lmm.3)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.3: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + POST1  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## lmm.0 4 1876 1897 -934 1868   
## lmm.3 5 1874 1901 -932 1864 3.49 1 0.062 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

lmm.3 <- update(lmm.0, .~. + PRE)   
anova(lmm.0,lmm.3)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.3: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + PRE  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)  
## lmm.0 4 1876 1897 -934 1868   
## lmm.3 5 1878 1904 -934 1868 0.1 1 0.75

lmm.3 <- update(lmm.0, .~. + POST1)   
anova(lmm.0,lmm.3)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.3: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + POST1  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## lmm.0 4 1876 1897 -934 1868   
## lmm.3 5 1874 1901 -932 1864 3.49 1 0.062 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

lmm.3 <- update(lmm.0, .~. + PRE\*POST1)   
anova(lmm.0,lmm.3)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.3: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + PRE + POST1 +   
## lmm.3: PRE:POST1  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## lmm.0 4 1876 1897 -934 1868   
## lmm.3 7 1871 1908 -928 1857 11.4 3 0.0096 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

lmm.4 <- update(lmm.0, .~. + AMSP)  
anova(lmm.0,lmm.4)

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

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.4: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + AMSP  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)  
## lmm.0 4 1876 1897 -934 1868   
## lmm.4 5 1877 1904 -934 1867 0.93 1 0.33

lmm.5 <- update(lmm.0, .~. + HRSD)  
anova(lmm.0,lmm.5)

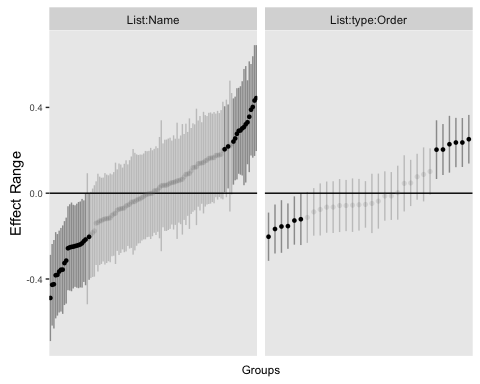
## refitting model(s) with ML (instead of REML)

## Data: DLM  
## Models:  
## lmm.0: log(Time) ~ (1 | List:Name) + (1 | List:type:Order)  
## lmm.5: log(Time) ~ (1 | List:Name) + (1 | List:type:Order) + HRSD  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)  
## lmm.0 4 1876 1897 -934 1868   
## lmm.5 5 1877 1904 -934 1867 0.74 1 0.39

We conclude that LEAYRS and the interaction PRE and POST1 are significant at a 0.05 significance level, while POST1 is significant at a 0.1 significance level.

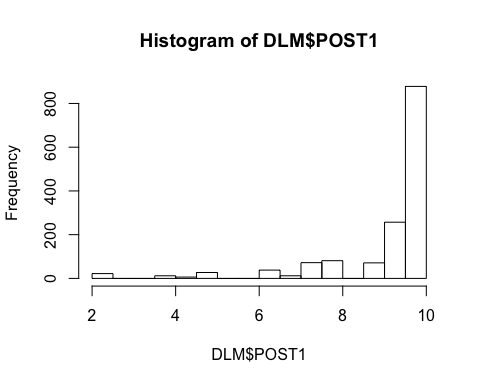
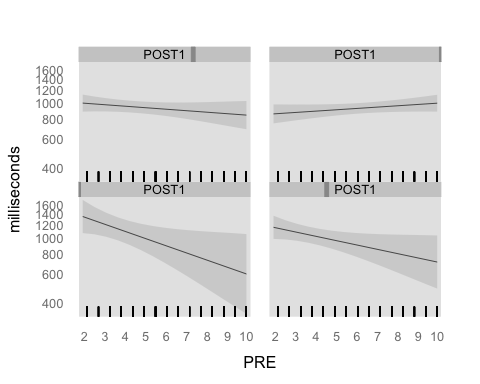
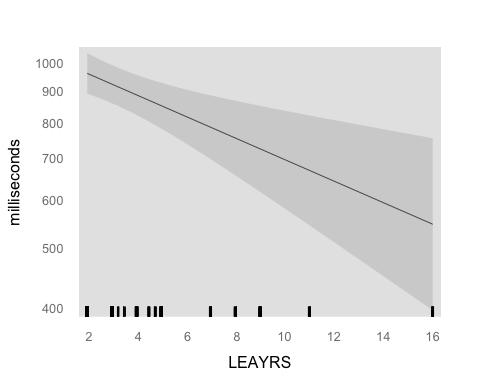
An automatic, step-backward procedure from the package lmerTest starting from a model that includes all second order fixed factor intereactions lead to a similar conclusion: the factors type, AMSP, HRSD are not significant while LEAYRS, PRE, POST1 and the interaction of these last two are significant.

The figures show random and the fixed effects. The random effects show that the different subjects imply a significantly different intercept. In the plot of the fixed effects, we back transformed the effects in linear scale and added 0.95-confidence level bands. It is evident the reduction in reaction time as learning years increase. As far as proficency is concerned, we see that at high values of POST1 (indicated by the dark grey vertical bar in the strip band) in the top two plots, the PRE has little impact, meaning that presence or absence of previous knoweledge of construction does not have an impact when the POST1 value is however high. On the contrary, for low values of POST1 the impact is considerable, in particular the subjects that achieved a value of the POST1 test lower than the value of the PRE test perform much better than those with a low value in both POST1 and PRE. The last plot shows the histogram of entries with respect to the POST1 value. The number of subjects with low POST1 values is however low, which is reflected in the much wider confidence bands of the preceeding plots.



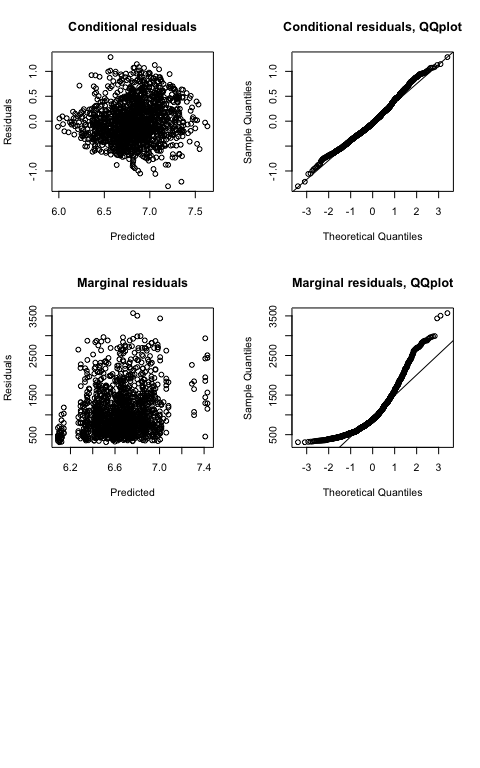
## Loading required package: effects

## Warning: package 'effects' was built under R version 3.2.4



## Diagnostic plots

par(mfrow=c(3,2))  
# plot(lm4,which=1:4)  
  
plot(fitted(lmm, type = "response"), residuals(lmm, type = "response"),  
 main = "Conditional residuals", xlab = "Predicted", ylab = "Residuals")  
  
res <- residuals(lmm, type = "response")  
qqnorm(res, main = "Conditional residuals, QQplot")  
qqline(res)  
  
lm.0 <- lm(log(Time) ~ ( type + LEAYRS + (POST1 + PRE)^2 + AMSP + HRSD), data = DLM)  
x <- model.matrix(lm.0)  
pred <- x %\*% fixef(lmm)  
res <- DLM$Time - pred  
plot(pred, res, main = "Marginal residuals", xlab = "Predicted", ylab = "Residuals")  
qqnorm(res, main = "Marginal residuals, QQplot")  
qqline(res)



The joint qqplot looks normal. The marginal looks less nice.

## Anova Table with Satterwhite

require(lmerTest)

## Loading required package: lmerTest

##   
## Attaching package: 'lmerTest'

## Det følgende objekt er maskeret fra 'package:lme4':  
##   
## lmer

## Det følgende objekt er maskeret fra 'package:stats':  
##   
## step

lmm <- lmer(log(Time) ~ type + LEAYRS + (PRE + POST1)^2 + AMSP + HRSD + POST1\*HRSD + (1 | List:Name) + (1 | List:type:Order), data = DLM, REML=FALSE)  
anova(lmm)

## Analysis of Variance Table of type III with Satterthwaite   
## approximation for degrees of freedom  
## Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)   
## type 0.021 0.021 1 31.5 0.12 0.7288   
## LEAYRS 1.792 1.792 1 110.8 10.32 0.0017 \*\*  
## PRE 1.005 1.005 1 138.1 5.79 0.0175 \*   
## POST1 1.680 1.680 1 120.8 9.67 0.0023 \*\*  
## AMSP 0.231 0.231 1 114.2 1.33 0.2514   
## HRSD 0.054 0.054 1 116.2 0.31 0.5771   
## PRE:POST1 1.143 1.143 1 136.0 6.59 0.0114 \*   
## POST1:HRSD 0.032 0.032 1 116.7 0.19 0.6661   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(lmm)

## Linear mixed model fit by maximum likelihood t-tests use  
## Satterthwaite approximations to degrees of freedom [lmerMod]  
## Formula:   
## log(Time) ~ type + LEAYRS + (PRE + POST1)^2 + AMSP + HRSD + POST1 \*   
## HRSD + (1 | List:Name) + (1 | List:type:Order)  
## Data: DLM  
##   
## AIC BIC logLik deviance df.resid   
## 1868.6 1932.2 -922.3 1844.6 1464   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.1318 -0.6654 -0.0973 0.5805 3.0847   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## List:Name (Intercept) 0.0532 0.231   
## List:type:Order (Intercept) 0.0206 0.143   
## Residual 0.1736 0.417   
## Number of obs: 1476, groups: List:Name, 118; List:type:Order, 32  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 7.78017 0.36703 127.60000 21.20 <2e-16 \*\*\*  
## typeMISMATCH -0.01938 0.05540 31.50000 -0.35 0.7288   
## LEAYRS -0.03994 0.01243 110.80000 -3.21 0.0017 \*\*   
## PRE -0.18200 0.07566 138.10000 -2.41 0.0175 \*   
## POST1 -0.11294 0.03631 120.80000 -3.11 0.0023 \*\*   
## AMSP 0.04331 0.03757 114.20000 1.15 0.2514   
## HRSD 0.07098 0.12693 116.20000 0.56 0.5771   
## PRE:POST1 0.02016 0.00786 136.00000 2.57 0.0114 \*   
## POST1:HRSD -0.00568 0.01312 116.70000 -0.43 0.6661   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) tMISMA LEAYRS PRE POST1 AMSP HRSD PRE:PO  
## typMISMATCH -0.075   
## LEAYRS -0.029 0.000   
## PRE -0.828 -0.004 -0.006   
## POST1 -0.892 -0.001 -0.047 0.830   
## AMSP -0.375 0.000 -0.104 0.069 0.008   
## HRSD -0.202 0.004 0.062 -0.019 0.209 -0.075   
## PRE:POST1 0.826 0.005 0.025 -0.986 -0.865 -0.071 0.036   
## POST1:HRSD 0.215 -0.004 -0.085 0.019 -0.208 0.041 -0.995 -0.039