Analysis I

"Aerosol emission is increased in professional singing'

Dirk Mürbe[1], Martin Kriegel[2], Julia Lange[2], Hansjörg Rotheudt[2], Mario Fleischer[1]

[1]Charité – Universitätsmedizin Berlin, Department of Audiology and Phoniatrics, Berlin, Germany,
[2]Technische Universität Berlin, Hermann-Rietschel-Institut, Berlin, Germany

Purpose

Linear mixed modeling of experiment I (breathing, speaking, singing).
Analysis is inspired by Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499

The advantages are
1) Using mixed modeling no (pre-)averaging is necessary
2) Condition & Gender (& Interfactorial effects) were modeled as fixed effects
3) ID was modeled as random effects (intercept and slope for Gender and Condition)
author: mario.fleischer@charite.de

```
rm(list = ls())
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
library(gridExtra)
library(lme4)
```

```
## Loading required package: Matrix
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```
library(redres) # https://github.com/goodekat/redres
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggResidpanel)
```

Loading data

```
# Set working directory according to your file hierarchy.
orig.data <- read.csv('Data.csv')
```

log-transform of emission rate and replacing zeros by NaN

```
PM.log <- orig.data$PM
PM.log[PM.log == 0] <- NA
orig.data$PM.log = log10(PM.log)
length(which(is.na(PM.log)==T)) # Number of NaN detected
```

```
## [1] 12
```

Selection of relevant data

```
exp.data.I <- dplyr::filter(orig.data,
                            Condition == 'breathing'
                            | Condition == 'speaking'
                            | Condition == 'singing'
                           )
```

Test 0: Deviation from normality and homoscedasticity?

```
I00.model <- lmer(PM~Condition*Gender
                  +(1+Condition|ID)
                  +(1+Gender|ID),
                 exp.data.I,
                 REML=FALSE,
                 control = lmerControl(optimizer ='optimx',
                                       optCtrl=list(method='nlminb'))
                 )
```

```
## Loading required namespace: optimx
```

```
## boundary (singular) fit: see ?isSingular
```

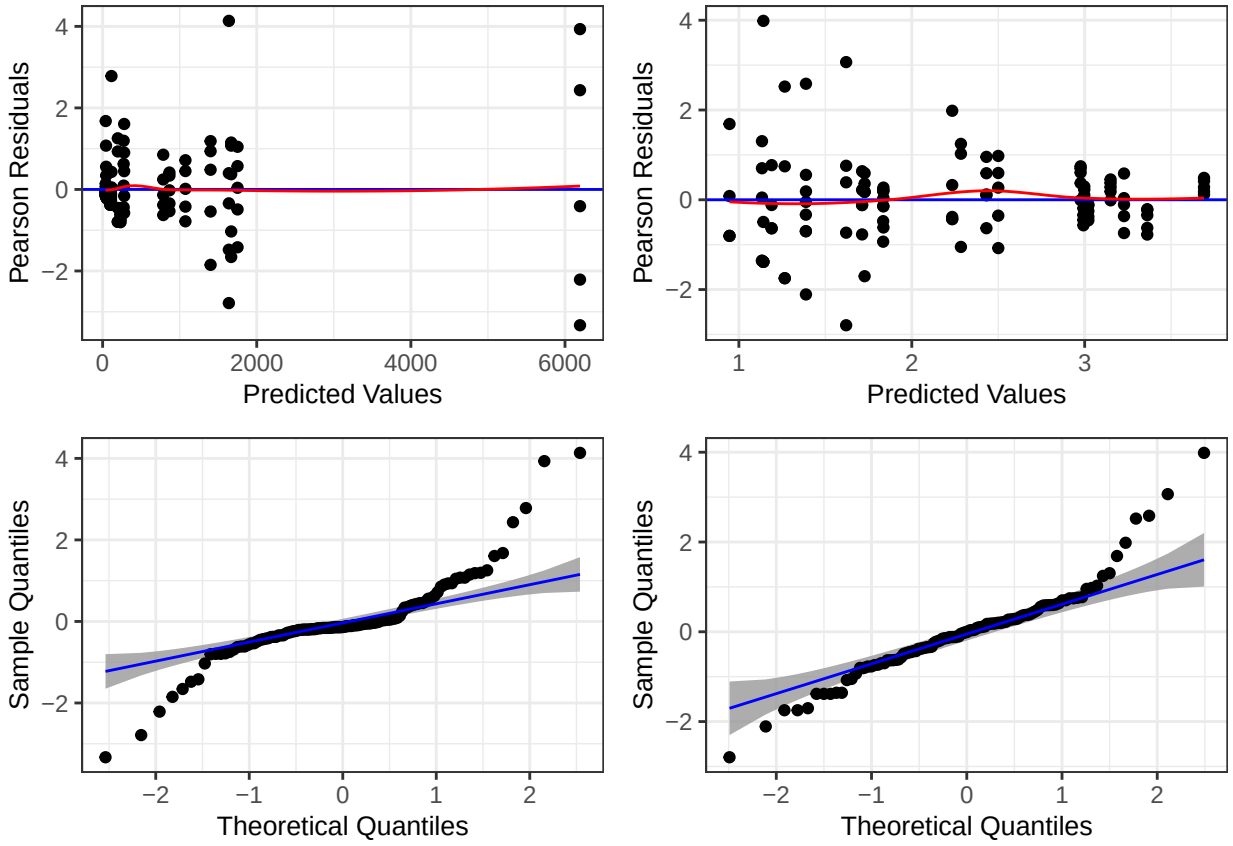
```
# (intermediate) results:
# 1) Deviation from normality and homoscedasticity was observed

I0.model <- lmer(PM.log~Condition*Gender
                  +(1+Condition|ID)
                  +(1+Gender|ID),
                 exp.data.I,
                 REML=FALSE,
                 control = lmerControl(optimizer ='optimx',
                                       optCtrl=list(method='nlminb'))
                 )
```

```
## boundary (singular) fit: see ?isSingular
```

Comparing model with and without log-transform of P_M

```
resid_compare(models = list(I00.model,
                            I0.model
                            ),
              plots = c("resid", "qq"),
              smoother = TRUE,
              qqbands = TRUE,
              title.opt = FALSE)
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```

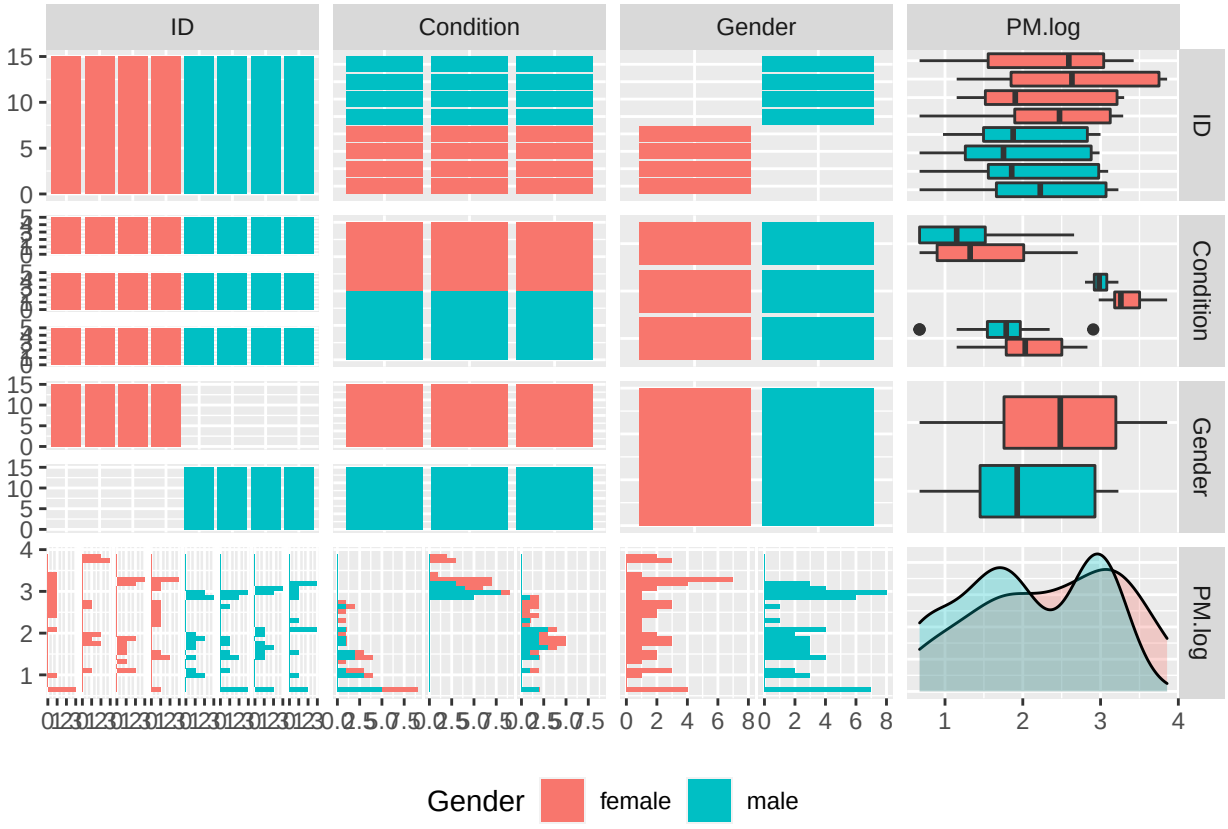


Pair plot of the data

```
exp.data.I %>% ggpairs(.,
                       mapping = ggplot2::aes(colour=Gender),
                       legend = 1,
                       column = c('ID','Condition','Gender','PM.log'),
                       lower = list(continuous = wrap("smooth", alpha = 0.3, size=0.1)),
                       diag = list(discrete="barDiag",
                                   continuous = wrap("densityDiag", alpha=0.3))
                       )+ theme(legend.position = "bottom")
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 7 rows containing non-finite values (stat_boxplot).

## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 7 rows containing non-finite values (stat_bin).
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 7 rows containing non-finite values (stat_bin).
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 7 rows containing non-finite values (stat_bin).
## Warning: Removed 7 rows containing non-finite values (stat_density).
```



Test 1: Is the interaction between Condition and Gender significant?

```
I1.model <- I0.model

I1.null <- lmer(PM.log~Condition+Gender
                  +(1+Condition|ID)
                  +(1+Gender|ID),
                  exp.data.I,
```

```
                  REML=FALSE,
                  control = lmerControl(optimizer ='optimx',
                                        optCtrl=list(method='nlminb'))
                )
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(I1.null, I1.model)
```

```
## Data: exp.data.I
## Models:
## I1.null: PM.log ~ Condition + Gender + (1 + Condition | ID) + (1 + Gender |
## I1.null:     ID)
## I1.model: PM.log ~ Condition * Gender + (1 + Condition | ID) + (1 + Gender |
## I1.model:     ID)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## I1.null    14 131.49 169.67 -51.742   103.48
## I1.model   16 135.41 179.05 -51.704   103.41 0.0768  2     0.9624
```

```
# (intermediate) results:
# 1) interaction between Condition and Gender is NOT significant
# 2) Condition:Gender is omitted for now
##############################################################################################
```

Test 2: Is Condition significant?

```
I2.model <- I1.null

I2.null <- lmer(PM.log~Gender
                +(1+Condition|ID)
                +(1+Gender|ID),
                exp.data.I,
                REML=FALSE,
                control = lmerControl(optimizer ='optimx',
                                      optCtrl=list(method='nlminb'))
                )
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(I2.null, I2.model)
```

```
## Data: exp.data.I
## Models:
## I2.null: PM.log ~ Gender + (1 + Condition | ID) + (1 + Gender | ID)
## I2.model: PM.log ~ Condition + Gender + (1 + Condition | ID) + (1 + Gender |
## I2.model:     ID)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## I2.null    12 165.28 198.01 -70.641   141.28
## I2.model   14 131.49 169.67 -51.742   103.48 37.797  2    6.2e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(I2.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
```

```
## Formula: PM.log ~ Condition + Gender + (1 + Condition | ID) + (1 + Gender |
##     ID)
##    Data: exp.data.I
## Control: lmerControl(optimizer = "optimx", optCtrl = list(method = "nlminb"))
##
##      AIC      BIC   logLik deviance df.resid
##    131.5    169.7    -51.7    103.5       99
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7757 -0.4807 -0.0064  0.3992  3.9641
##
## Random effects:
##  Groups   Name              Variance Std.Dev. Corr
##  ID       (Intercept)       0.07513  0.2741
##           Conditionsinging  0.06588  0.2567   -1.00
##           Conditionspeaking 0.51619  0.7185   -1.00  1.00
##  ID.1     (Intercept)       0.04476  0.2116
##           Gendermale        0.04476  0.2116   -1.00
##  Residual                   0.11490  0.3390
## Number of obs: 113, groups:  ID, 8
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)          1.5760     0.1592   9.900
## Conditionsinging     1.7740     0.1211  14.651
## Conditionspeaking    0.5230     0.2664   1.963
## Gendermale          -0.3453     0.1246  -2.772
##
## Correlation of Fixed Effects:
##             (Intr) Cndtnsn Cndtnsp
## Condtnsngng -0.644
## Condtnspkng -0.666  0.825
## Gendermale  -0.680  0.008   0.004
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
# (intermediate) results:
# 1) Condition IS significant. Condition affected log10(PM) (Xi^2(2)=37.797,p=6.2e-9),
#       increasing it by a factor of 0.5230 P/s+-0.2664 (standard errors)
#       from breathing to speaking and by a factor of 1.7740+-0.1211 from breathing to
#       singing
##############################################################################################
```

Test 3: Is Gender significant?

```
I3.model <- I1.null

I3.null <- lmer(PM.log~Condition
                +(1+Condition|ID)
                +(1+Gender|ID),
                exp.data.I,
```

```
                REML=FALSE,
                control = lmerControl(optimizer ='optimx',
                                      optCtrl=list(method='nlminb'))
               )
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(I3.null, I3.model)
```

```
## Data: exp.data.I
## Models:
## I3.null: PM.log ~ Condition + (1 + Condition | ID) + (1 + Gender | ID)
## I3.model: PM.log ~ Condition + Gender + (1 + Condition | ID) + (1 + Gender |
## I3.model:     ID)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## I3.null    13 133.79 169.24 -53.894   107.79
## I3.model   14 131.49 169.67 -51.742   103.48 4.3035  1    0.03803 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# (intermediate) results:
# 1) Gender IS significant. Gender affected log10(PM) (Xi^2(1)=4.3035,p=0.03803),
#       lowering it by a factor of -0.3453 P/s+-0.1246 (standard errors)
#       from female to male
#############################################################################################
```

By-subject analysis

```
coef(I3.model)
```

```
## $ID
##    (Intercept) Conditionsinging Conditionspeaking Gendermale
## S1   0.9427102         2.070445          1.3528669 -0.2347422
## S2   2.4091803         1.383835         -0.5690541 -0.6500869
## S3   1.6412443         1.743387          0.4373846 -0.1351186
## S4   1.1836732         1.957624          1.0370665 -0.3612540
## S5   1.4754273         1.821023          0.6547005 -0.3452990
## S6   2.3424326         1.415087         -0.4815761 -0.3452997
## S7   1.6277395         1.749710          0.4550837 -0.3453004
## S8   0.9852099         2.050546          1.2971678 -0.3453025
##
## attr(,"class")
## [1] "coef.mer"
```

```
# 1) all subjects (except for S2 and S6 - see Fig. 2 in manuscript)
#     show an increase of emitted Particles from breathing to speaking
# 2) all subjects
#     show an increase of emitted Particles from breathing to singing
# 3) all subjects show, that females emitted more particles than males
# 4) the highest ratio was shown by S2
```