practical_exercise_2, Methods 3, 2021, autumn semester

Sirid Wihlborg

29/09/21

Assignment 1: Using mixed effects modelling to model hierarchical data

In this assignment we will be investigating the *politeness* dataset of Winter and Grawunder (2012) and apply basic methods of multilevel modelling.

Dataset

The dataset has been shared on GitHub, so make sure that the csv-file is on your current path. Otherwise you can supply the full path.

```
politeness <- read.csv('politeness.csv') ## read in data
politeness <- na.omit(politeness)
pacman::p_load(tidyverse, lme4, car,lmerTest)
```

Exercises and objectives

The objectives of the exercises of this assignment are:
1) Learning to recognize hierarchical structures within datasets and describing them
2) Creating simple multilevel models and assessing their fitness
3) Write up a report about the findings of the study

REMEMBER: In your report, make sure to include code that can reproduce the answers requested in the exercises below
REMEMBER: This assignment will be part of your final portfolio

Exercise 1 - describing the dataset and making some initial plots

1) Describe the dataset, such that someone who happened upon this dataset could understand the variables and what they contain
    i. Also consider whether any of the variables in *politeness* should be encoded as factors or have the factor encoding removed. Hint: `?factor`

Explaining the data-set

The experiment that the data relies set out to investigate whether our pitch changes depending on if we are in a formal or informal setting. The experiment was done in Korea. Each participant went through two conditions (column: attitude) either an informal or formal. They had to read out loud a pre-printed sentence and this recording was analysed in terms of pitch so the variable contains the mean pitch in Hz pr sentence (column: f0mn). Besides these variable we have a variable expressing gender (F = Female, M = Male), a variable where the scenario is given (scenario: an integer 1:7).

```
politeness$gender <- as.factor(politeness$gender)
politeness$scenario <- as.factor(politeness$scenario)
```

2) Create a new data frame that just contains the subject *F1* and run two linear models; one that expresses *f0mn* as dependent on *scenario* as an integer; and one that expresses *f0mn* as dependent on *scenario* encoded as a factor

```
sub_poli <- politeness[which(politeness$subject == "F1"),]

lm1 <- lm(f0mn~as.factor(scenario), data = sub_poli)
lm2 <- lm(f0mn~as.integer(scenario), data = sub_poli)
```

```
i. Include the model matrices, $X$ from the General Linear Model, for these two models in your report ar
ii. Which coding of _scenario_, as a factor or not, is more fitting?
```

```
print(model.matrix(lm1)) # print design matrix for factor model
```

```
##    (Intercept) as.factor(scenario)2 as.factor(scenario)3 as.factor(scenario)4
## 1            1                    0                    0                    0
## 2            1                    0                    0                    0
## 3            1                    1                    0                    0
## 4            1                    1                    0                    0
## 5            1                    0                    1                    0
## 6            1                    0                    1                    0
## 7            1                    0                    0                    1
## 8            1                    0                    0                    1
## 9            1                    0                    0                    0
## 10           1                    0                    0                    0
## 11           1                    0                    0                    0
## 12           1                    0                    0                    0
## 13           1                    0                    0                    0
## 14           1                    0                    0                    0
##    as.factor(scenario)5 as.factor(scenario)6 as.factor(scenario)7
## 1                     0                    0                    0
## 2                     0                    0                    0
## 3                     0                    0                    0
## 4                     0                    0                    0
## 5                     0                    0                    0
## 6                     0                    0                    0
## 7                     0                    0                    0
## 8                     0                    0                    0
## 9                     1                    0                    0
## 10                    1                    0                    0
## 11                    0                    1                    0
## 12                    0                    1                    0
## 13                    0                    0                    1
## 14                    0                    0                    1
```

```
## attr(,"assign")
## [1] 0 1 1 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$`as.factor(scenario)`
## [1] "contr.treatment"
```

```
print(model.matrix(lm2)) # print design matrix for integer model
```

```
##    (Intercept) as.integer(scenario)
## 1            1                    1
## 2            1                    1
## 3            1                    2
## 4            1                    2
## 5            1                    3
## 6            1                    3
## 7            1                    4
## 8            1                    4
## 9            1                    5
## 10           1                    5
## 11           1                    6
## 12           1                    6
## 13           1                    7
## 14           1                    7
## attr(,"assign")
## [1] 0 1
```

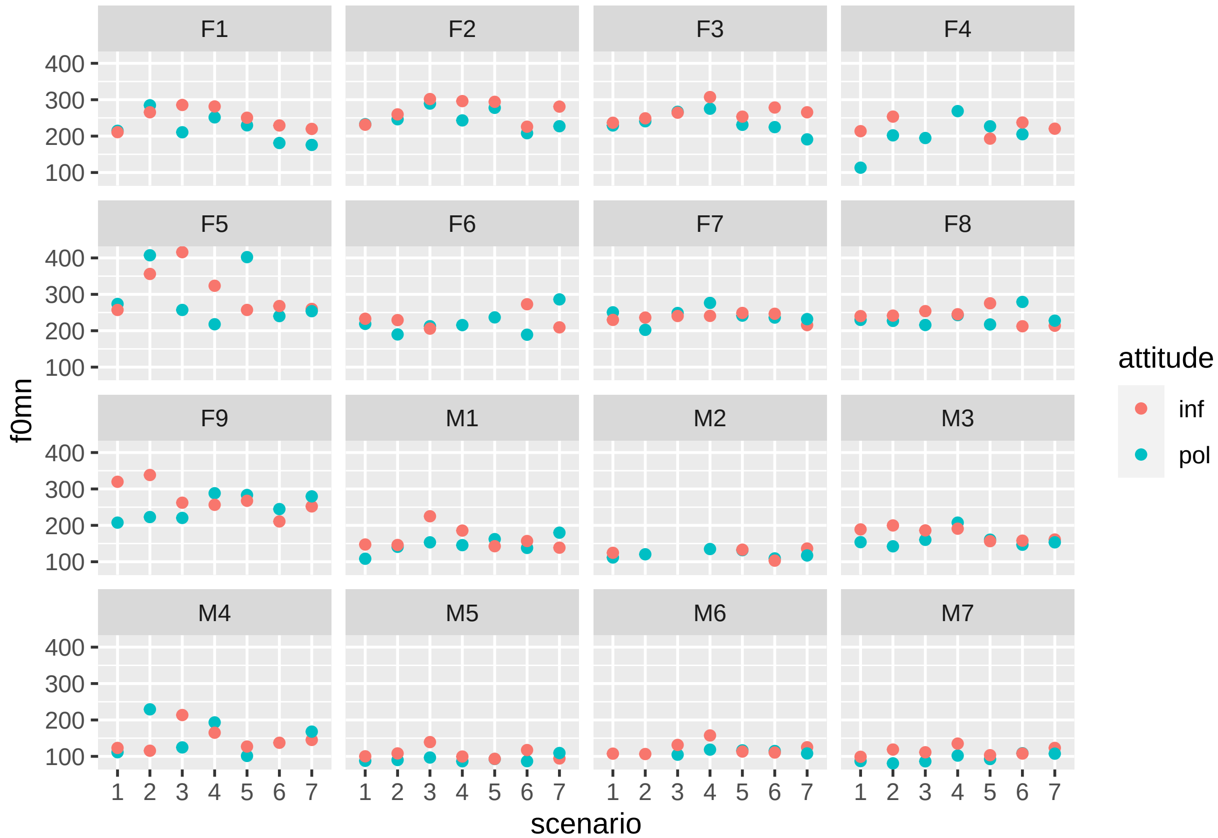
Having scenario as an integer will make the model mistakenly interpret each scenario as 7 numerical values, where there are relationship between the number which also implies that we hypothetically could "predict" the pitch of a scenario 8 from the pitch in scenario 7, which doesn't make sense. Therefor it makes sense to treat scenario as a factor, having 7 different independent scenario.

I can't really explain why the design matrices look the way they do, maybe you could elaborate on this in class :)

3) Make a plot that includes a subplot for each subject that has *scenario* on the x-axis and *f0mn* on the y-axis and where points are colour coded according to *attitude*

    i. Describe the differences between subjects

```
ggplot(data = politeness, aes(x = scenario, y = f0mn, color = attitude)) +
  geom_point() +
  facet_wrap(~subject)
```



We see that the different participants have clearly different baselines meaning that they speak with generally different pitch, which makes good sense.

Exercise 2 - comparison of models

For this part, make sure to have `lme4` installed.
You can install it using `install.packages("lme4")` and load it using `library(lme4)`
`lmer` is used for multilevel modelling

```
#mixed.model <- lmer(formula=..., data=...)
#example.formula <- formula(dep.variable ~ first.level.variable + (1 | second.level.variable))
```

1) Build four models and do some comparisons
    i. a single level model that models *f0mn* as dependent on *gender*
    ii. a two-level model that adds a second level on top of i. where unique intercepts are modelled for each *scenario*
    iii. a two-level model that only has *subject* as an intercept
    iv. a two-level model that models intercepts for both *scenario* and *subject*

```
m1 <- lm(f0mn ~ gender, data = politeness)
m2 <- lmerTest::lmer(f0mn ~ gender + (1|scenario), data = politeness, REML = FALSE)
m3 <- lmerTest::lmer(f0mn ~ gender + (1|subject), data = politeness, REML = FALSE)
m4 <- lmerTest::lmer(f0mn ~ gender + (1|scenario) + (1|subject), data = politeness, REML = FALSE)
```

```
v. which of the models has the lowest residual standard deviation, also compare the Akaike Information C
```

```r
# Finding residual standard deviation
tibble(sigma(m1), sigma(m2), sigma(m3), sigma(m4))
```

```
## # A tibble: 1 x 4
##   `sigma(m1)` `sigma(m2)` `sigma(m3)` `sigma(m4)`
##         <dbl>       <dbl>       <dbl>       <dbl>
## 1        39.5        38.4        32.0        30.7
```

```r
# Finding AIC
AIC(m1, m2, m3, m4)
```

```
##    df      AIC
## m1  3 2163.971
## m2  4 2162.257
## m3  4 2112.048
## m4  5 2105.176
```

The fourth model: a two-level model that models intercepts for both *scenario* and *subject*. This model has the lowest AIC value and lowest residual standard deviation.

`vi. which of the second-level effects explains the most variance?`

```r
summary(m4)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender + (1 | scenario) + (1 | subject)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##   2105.2   2122.0  -1047.6   2095.2      207
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0357 -0.5384 -0.1177  0.4346  3.7808
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 516.19   22.720
##  scenario (Intercept)  89.36    9.453
##  Residual             940.25   30.664
## Number of obs: 212, groups:  subject, 16; scenario, 7
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  246.778      8.829   19.248  27.952  < 2e-16 ***
## genderM     -115.186     12.223   16.011  -9.424 6.19e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## genderM -0.604
```

2) Why is our single-level model bad?

Because we have some systemacy in our error term (like subject and gender) so putting these as random effects drastically helps our model to "make sense" of the noise.

```
i. create a new data frame that has three variables, _subject_, _gender_ and _f0mn_, where _f0mn_ is the
```

```
politeness2 <- politeness %>%
  group_by(subject, gender) %>%
  summarise(mean_pitch = mean(f0mn))

head(politeness2)
```

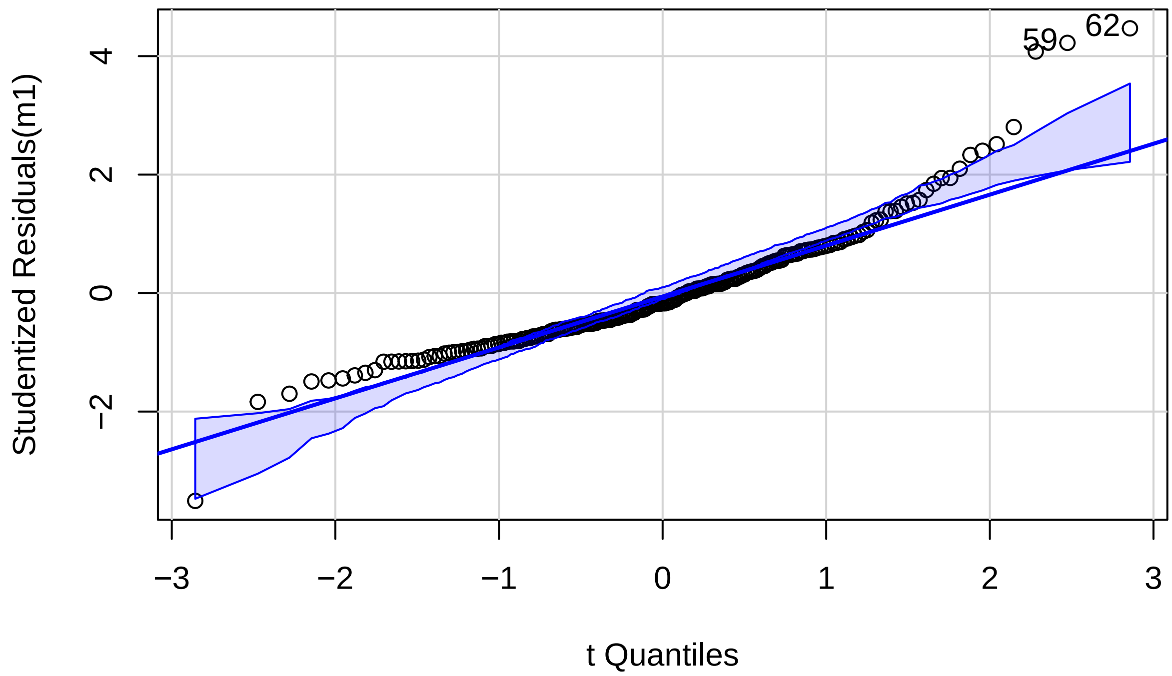
```
## # A tibble: 6 x 3
## # Groups:   subject [6]
##   subject gender mean_pitch
##   <chr>   <fct>       <dbl>
## 1 F1      F            235.
## 2 F2      F            258.
## 3 F3      F            251.
## 4 F4      F            212.
## 5 F5      F            299.
## 6 F6      F            225.
```

```
ii. build a single-level model that models _f0mn_ as dependent on _gender_ using this new dataset
```

```
m5 <- lm(mean_pitch ~ gender, data = politeness2)
```

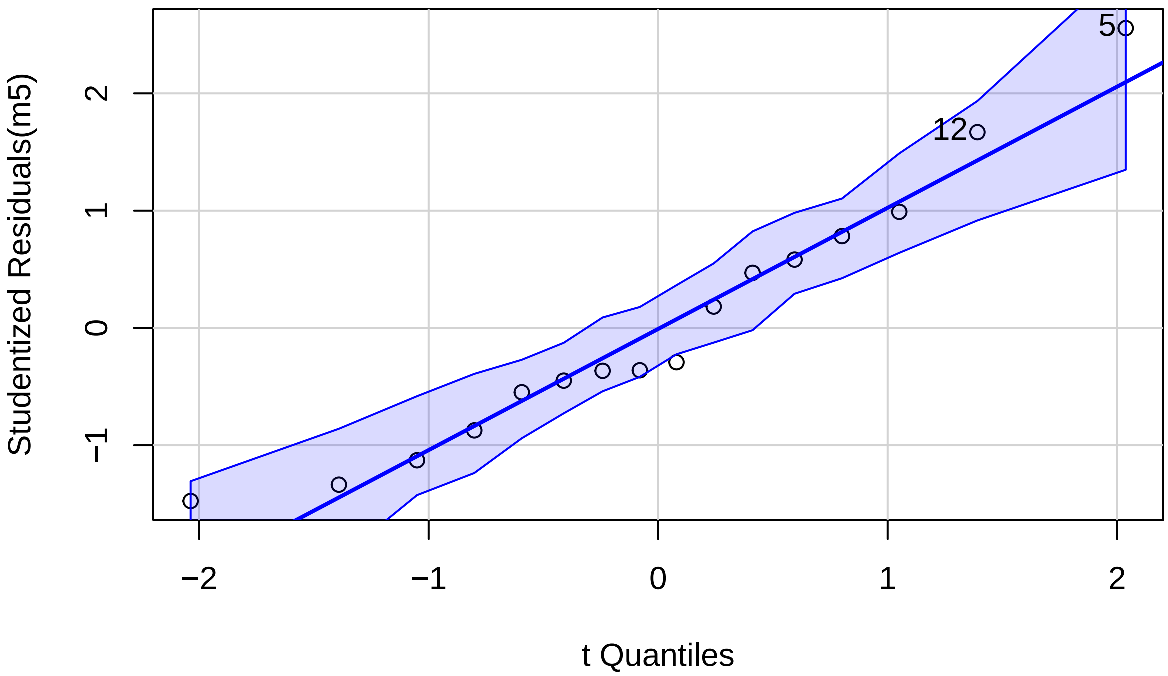
```
iii. make Quantile-Quantile plots, comparing theoretical quantiles to the sample quantiles) using ‘qqno
```

```
qqPlot(m1) # Plot using all pitch scores pr. participant
```




```
## 59 62
## 56 59
```

```
qqPlot(m5) # Plot using the averaged pitch score pr. participant
```

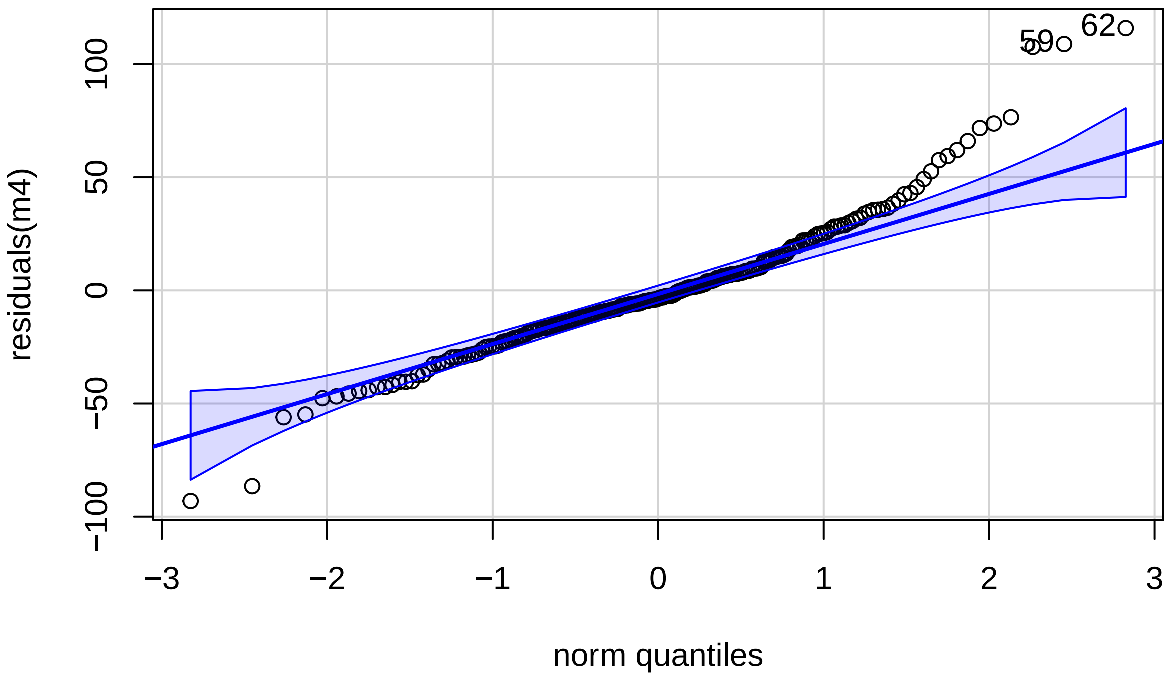



```
## [1]  5 12
```

We see that the qqplot on the averaged data is in fact more close to being normally distributed. However, this might be an illusion since we don't have a lot of data points so we are for sure more insecure on judging residual normality.

```
iv. Also make a quantile-quantile plot for the residuals of the  multilevel model with two intercepts. I
```

```
qqPlot(residuals(m4))
```

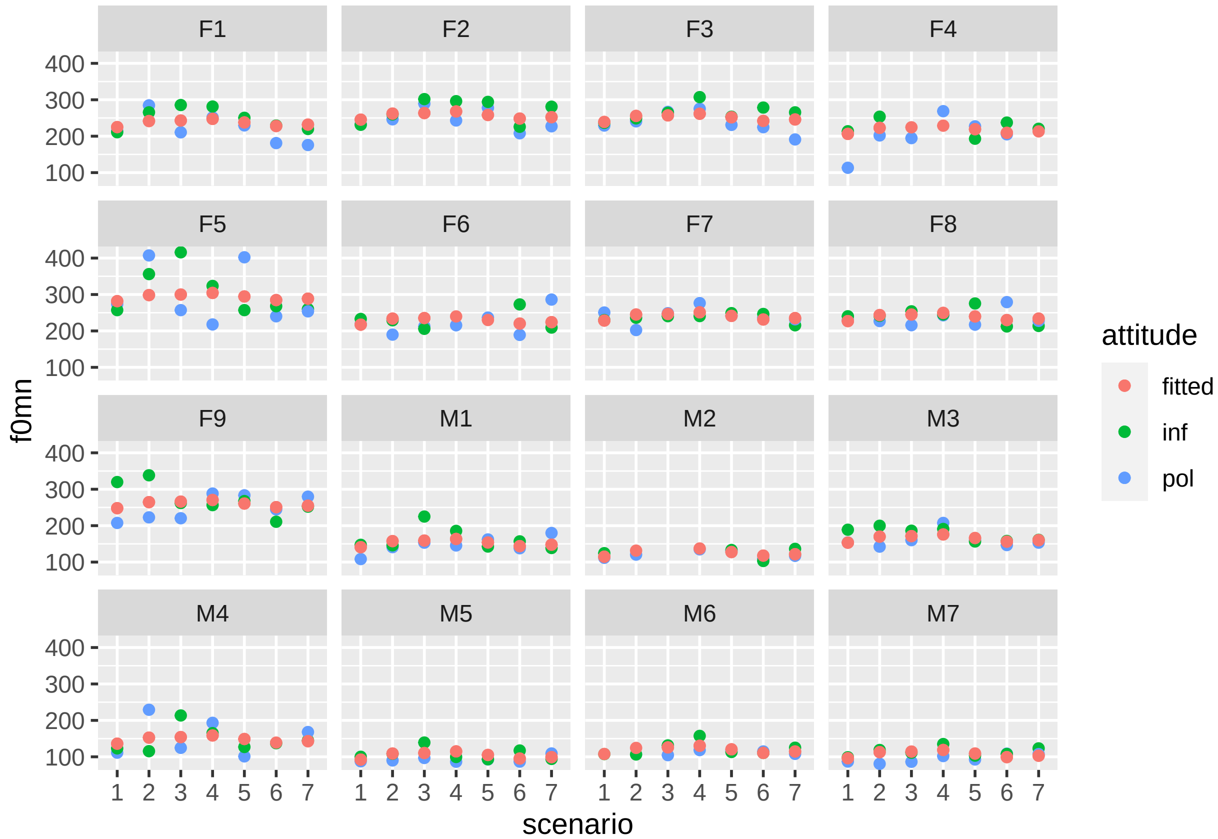



```
## 62 59
## 59 56
```

3) Plotting the two-intercepts model

   i. Create a plot for each subject, (similar to part 3 in Exercise 1), this time also indicating the fitted value for each of the subjects for each for the scenarios (hint use `fixef` to get the "grand effects" for each gender and `ranef` to get the subject- and scenario-specific effects)

```
ggplot(data = politeness, aes(x = scenario, y = f0mn, color = attitude)) +
  geom_point() +
  geom_point(aes(x = scenario, y = fitted(m4), color = "fitted")) + # adding the fitted values pr. subj
  facet_wrap(~subject)
```



Exercise 3 - now with attitude

1) Carry on with the model with the two unique intercepts fitted (*scenario* and *subject*).
    i. now build a model that has *attitude* as a main effect besides *gender*

```
m6 <- lmerTest::lmer(f0mn ~ gender + attitude + (1 | scenario) + (1 | subject), data = politeness, REML
```

```
ii. make a separate model that besides the main effects of _attitude_ and _gender_ also include their i
```

```
m7 <- lmerTest::lmer(f0mn ~ gender * attitude + (1 | scenario) + (1 | subject), data = politeness, REML
```

```
iii. describe what the interaction term in the model says about Korean men's pitch when they are polite
```

```
fixef(m7)
```

```
##         (Intercept)             genderM         attitudepol genderM:attitudepol
##          255.632258         -118.250699          -17.198433            5.562949
```

We see that the p-value is way above the threshold (.05) indicating no significant interaction effect on gender and pitch difference. However, had we assumed that the p-value was below .05 we could have concluded: The interaction term is positive meaning that men generally lower their pitch *less* when changing to an informal setting.

2) Compare the three models (1. gender as a main effect; 2. gender and attitude as main effects; 3. gender and attitude as main effects and the interaction between them. For all three models model unique intercepts for *subject* and *scenario*) using residual variance, residual standard deviation and AIC.

```
# m4: f0mn ~ gender + (1 | scenario) + (1 | subject)
# m6:  f0mn ~ gender + attitude + (1 | scenario) + (1 | subject)
# m7: f0mn ~ gender * attitude + (1 | scenario) + (1 | subject)

# Finding sum of residual variance
tibble(sum(residuals(m4)^2),
       sum(residuals(m6)^2),
       sum(residuals(m7)^2))
```

```
## # A tibble: 1 x 3
##   `sum(residuals(m4)^2)` `sum(residuals(m6)^2)` `sum(residuals(m7)^2)`
##                    <dbl>                  <dbl>                  <dbl>
## 1                181913.                169681.                169306.
```

```
# Finding residual standard deviation
tibble(sigma(m4), sigma(m6), sigma(m7))
```

```
## # A tibble: 1 x 3
##   `sigma(m4)` `sigma(m6)` `sigma(m7)`
##         <dbl>       <dbl>       <dbl>
## 1        30.7        29.6        29.6
```

```
# Finding AIC
AIC(m4, m6, m7)
```

```
##    df      AIC
## m4  5 2105.176
## m6  6 2094.489
## m7  7 2096.034
```

3) Choose the model that you think describe the data the best - and write a short report on the main findings based on this model. At least include the following:

i. describe what the dataset consists of

ii. what can you conclude about the effect of gender and attitude on pitch (if anything)?

iii. motivate why you would include separate intercepts for subjects and scenarios (if you think they should be included)

iv. describe the variance components of the second level (if any)

v. include a Quantile-Quantile plot of your chosen model

This paper uses data fronm an experiment by Bodo Winter (2012) set out to investigate whether our pitch changes depending on if we are in a formal or informal setting. The experiment was done in Korea. Each participant went through two conditions (column: attitude) either an informal or formal. They had to read out loud a pre-printed sentence and this recording was analysed in terms of pitch so the variable contains the mean pitch in Hz pr sentence (column: f0mn).

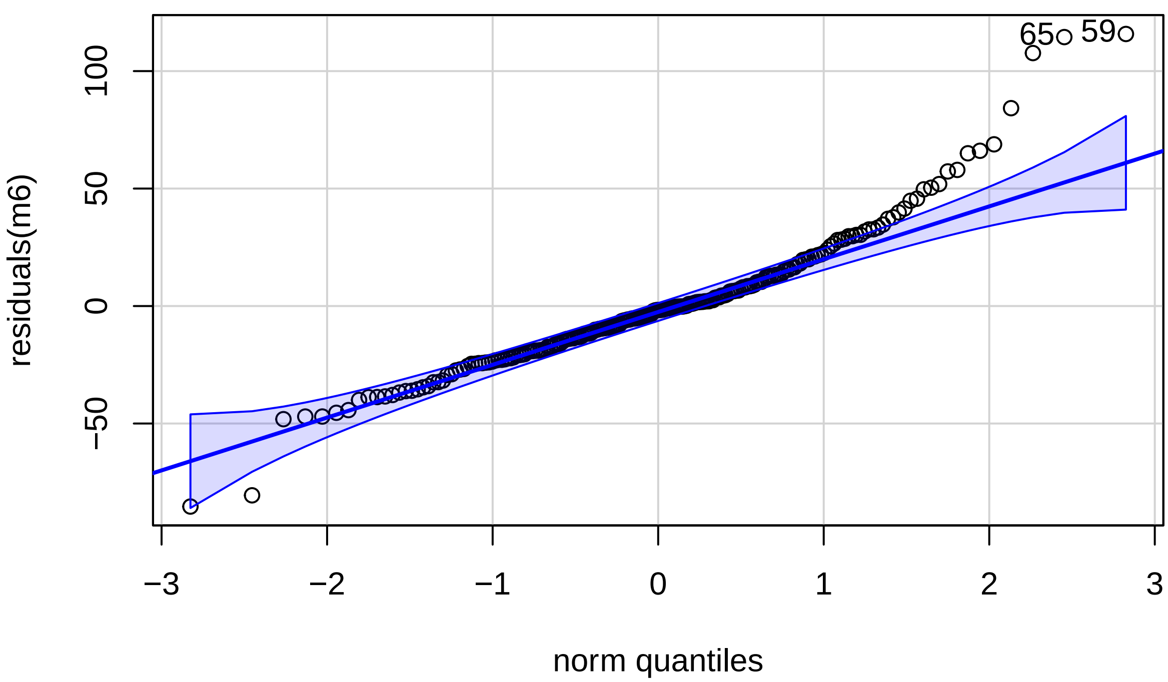
Generally, when choosing my model I would probably never primarily choose the model given what's describing the data best, but rather what makes sense given my hypothesis and assumption. So, for example if my hypothesis explicitly investigated the effect of the formality in a given context on pitch, I would for sure want "attitude" (the formal/informal condition) as a fixed effect, thereby excluding m4 (f0mn ~ gender + (1 | scenario) + (1 | subject)).

Given that including the interaction is both yielding insignificance and raising the AIC value, I've decided to go with the two-level model that predicts pitch (f0mn) with gender and attitude as fixed effects and subject and scenario as random effects, but including no interaction: f0mn ~ gender + attitude + (1 | scenario) + (1 | subject).

Subject and scenario has been set as random intercepts as I expect different baseline pitch for each participant and different effect on pitch for each scenario. We see that the subject random effect in fact explains quite a substantial part of the total variance, whereas scenario explains a lesser part.

Results: By setting the alpha level to 5 %, I can conclude that men have a significantly lower pitch (men: 139 Hz) than women (woman: 254 Hz) ($p < .05$). I can also conclude that in a formal context there was found a significantly lower pitch by 14.8 Hz ($p < .05$).

```
qqPlot(residuals(m6))
```




```
## 59 65
## 56 62
```

The residual qqplot reveals that the residuals are a bit right-skewed, but I would accept it.