practical_exercise_2, Methods 3, 2021, autumn semester

Aleksander M. Wael

29/09/2021

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

# Omit na
politeness <- na.omit(politeness)
```

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

Data comes from an experiment in which researches are interested in differences of voice pitch in formal and informal settings. The dataset contains 224 observations with 7 variables to describe the data. These variables are labeled: - Subject: The anonymized participant ID. - Gender: The gender of the participant, either F (female) or M (male). - Scenario: Categorized as 7 different scenarios of dialogue, labeled 1-7. - Attitude: The attitude of the scenario, either formal or informal. - Total duration: Duration of the scenario.

- f0mn: The mean pitch of the participants voice in a scenario. - hiss_count: How many hisses the subject utters during the scenario.

i. Also consider whether any of the variables in *politeness* should be encoded as factors or have the factor encoding removed. Hint: `?factor`

```
# First 4 variables are
politeness$subject <- as.factor(politeness$subject)
politeness$gender <- as.factor(politeness$gender)
politeness$scenario <- as.factor(politeness$scenario)
politeness$attitude <- as.factor(politeness$attitude)
politeness$total_duration <- as.numeric(politeness$total_duration)
politeness$f0mn <- as.numeric(politeness$f0mn)
politeness$hiss_count <- as.numeric(politeness$hiss_count)

summary(lm(total_duration ~ attitude, data = politeness))
```

```
##
## Call:
## lm(formula = total_duration ~ attitude, data = politeness)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -22.67 -17.99 -11.75  16.47  76.96
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  25.1444     2.4072  10.446   <2e-16 ***
## attitudepol  -0.7275     3.3883  -0.215     0.83
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.67 on 210 degrees of freedom
## Multiple R-squared:  0.0002195,  Adjusted R-squared:  -0.004541
## F-statistic: 0.0461 on 1 and 210 DF,  p-value: 0.8302
```

I encode the first 4 variables as factors because i want to treat them as

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

```
# Make subset of data to only include F1
subject_integer_df <- politeness %>%
  filter(subject == "F1")

subject_factor_df <- politeness %>%
  filter(subject == "F1")

# Treat subject as integer
# I already changed subject to factor previously, but will write the code again to practice
subject_integer_df$scenario <- as.integer(subject_integer_df$scenario)
```

```r
subject_factor_df$scenario <- as.factor(subject_factor_df$scenario)

# Make model with the different encodings
model_as_integer <- lm(f0mn ~ scenario, data = subject_integer_df)
model_as_factor <- lm(f0mn ~ scenario, data = subject_factor_df)
```

i. Include the model matrices, X from the General Linear Model, for these two models in your report and describe the different interpretations of *scenario* that these entail

```r
# Include model matrix
model.matrix(model_as_integer)
```

```
##    (Intercept) 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
```

```r
model.matrix(model_as_factor)
```

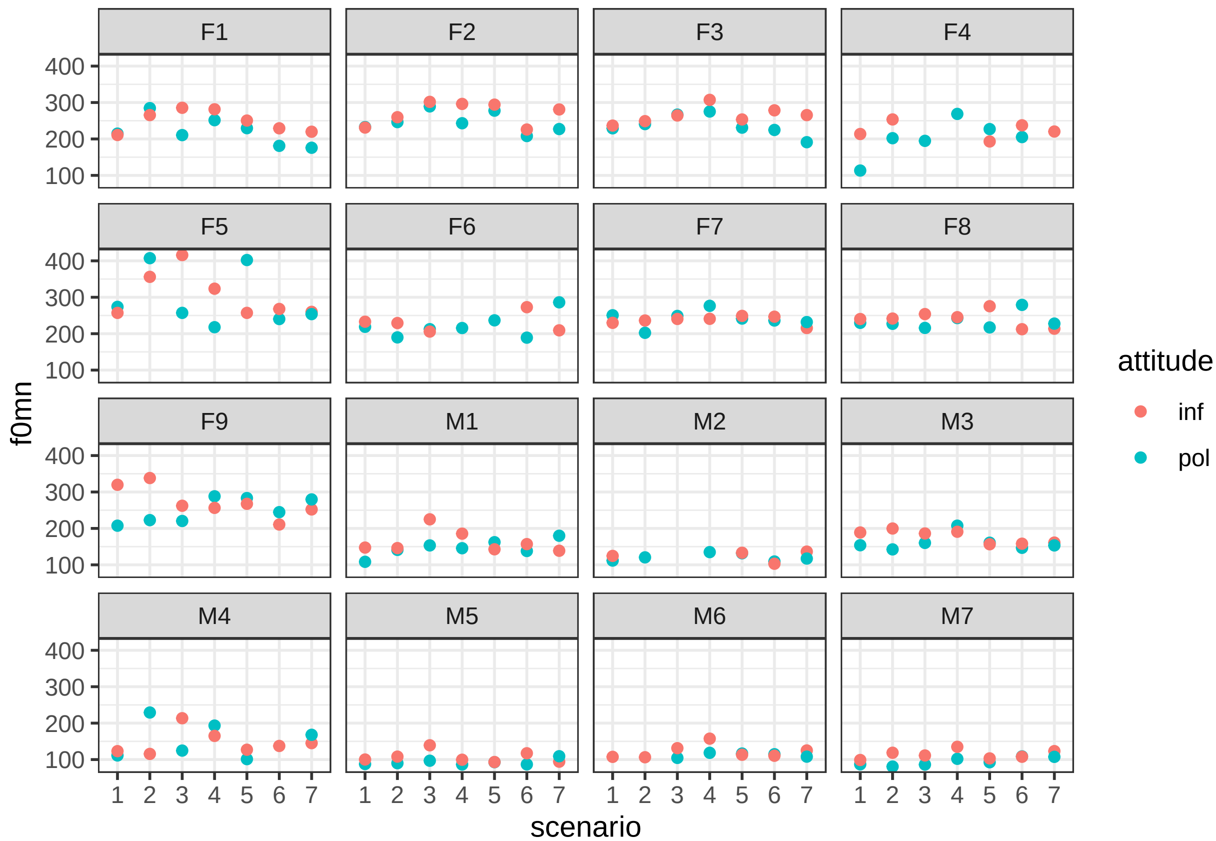
```
##    (Intercept) scenario2 scenario3 scenario4 scenario5 scenario6 scenario7
## 1            1         0         0         0         0         0         0
## 2            1         0         0         0         0         0         0
## 3            1         1         0         0         0         0         0
## 4            1         1         0         0         0         0         0
## 5            1         0         1         0         0         0         0
## 6            1         0         1         0         0         0         0
## 7            1         0         0         1         0         0         0
## 8            1         0         0         1         0         0         0
## 9            1         0         0         0         1         0         0
## 10           1         0         0         0         1         0         0
## 11           1         0         0         0         0         1         0
## 12           1         0         0         0         0         1         0
## 13           1         0         0         0         0         0         1
## 14           1         0         0         0         0         0         1
## attr(,"assign")
## [1] 0 1 1 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$scenario
## [1] "contr.treatment"
```

ii. Which coding of *scenario*, as a factor or not, is more fitting?

We are interested in treating each scenario as separate entities that aren't ranked based on their namings, i.e. scenario 7 is not larger than scenario 5; it is simply a different condition to compare to.

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*

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



i. Describe the differences between subjects

Some subjects have greater differences between attitude conditions, and the baseline f0mn also appears to be varying.

Exercise 2 - comparison of models

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

```r
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*

```r
m1 <- lm(f0mn ~ gender, data = politeness)
summary(m1)
```

```
##
## Call:
## lm(formula = f0mn ~ gender, data = politeness)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -134.283  -24.928   -6.783   20.517  168.217
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  247.583      3.588   69.01   <2e-16 ***
## genderM     -115.821      5.476  -21.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.46 on 210 degrees of freedom
## Multiple R-squared:  0.6806, Adjusted R-squared:  0.679
## F-statistic: 447.4 on 1 and 210 DF,  p-value: < 2.2e-16
```

```
ii. a two-level model that adds a second level on top of i. where unique intercepts are modelled for ea
```

```r
m2 <- lmerTest::lmer(f0mn ~ gender + (1|scenario), data = politeness, REML = FALSE)
summary(m2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender + (1 | scenario)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##   2162.3   2175.7  -1077.1   2154.3      208
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2617 -0.6192 -0.1537  0.4899  4.2318
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  scenario (Intercept)   71.82   8.475
##  Residual             1471.08  38.355
## Number of obs: 212, groups:  scenario, 7
```

```
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  247.768      4.735   11.793   52.32  2.5e-15 ***
## genderM     -115.870      5.324  205.219  -21.76  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## genderM -0.483
```

```
summary(lmerTest::lmer(f0mn ~ gender + (1|scenario), data = politeness, REML = FALSE))
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender + (1 | scenario)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##   2162.3   2175.7  -1077.1   2154.3      208
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2617 -0.6192 -0.1537  0.4899  4.2318
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  scenario (Intercept)   71.82   8.475
##  Residual             1471.08  38.355
## Number of obs: 212, groups:  scenario, 7
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  247.768      4.735   11.793   52.32  2.5e-15 ***
## genderM     -115.870      5.324  205.219  -21.76  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## genderM -0.483
```

iii. a two-level model that only has _subject_ as an intercept

```
m3 <- lmerTest::lmer(f0mn ~ gender + (1|subject), data = politeness, REML = FALSE)
summary(m3)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender + (1 | subject)
##    Data: politeness
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   2112.0   2125.5  -1052.0   2104.0      208
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2405 -0.5471 -0.1431  0.4360  3.8443
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept)  511.2   22.61
##  Residual             1026.7   32.04
## Number of obs: 212, groups:  subject, 16
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  246.547      8.083   15.984  30.501 1.36e-15 ***
## genderM     -115.193     12.239   16.076  -9.412 6.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr)
## genderM -0.660
```

iv. a two-level model that models intercepts for both _scenario_ and _subject_

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

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
# AIC values
tibble(AIC(m1), AIC(m2), AIC(m3), AIC(m4))
```

```
## # A tibble: 1 x 4
##   `AIC(m1)` `AIC(m2)` `AIC(m3)` `AIC(m4)`
##       <dbl>     <dbl>     <dbl>     <dbl>
## 1     2164.     2162.     2112.     2105.
```

```r
anova(m2, m1, m3, m4)
```

```
## Data: politeness
## Models:
## m1: f0mn ~ gender
## m2: f0mn ~ gender + (1 | scenario)
## m3: f0mn ~ gender + (1 | subject)
## m4: f0mn ~ gender + (1 | scenario) + (1 | subject)
##    npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m1    3 2164.0 2174.0 -1079.0   2158.0
## m2    4 2162.3 2175.7 -1077.1   2154.3  3.7136  1   0.053969 .
## m3    4 2112.1 2125.5 -1052.0   2104.1 50.2095  0
## m4    5 2105.2 2122.0 -1047.6   2095.2  8.8725  1   0.002895 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

We see by comparing residual standard deviation and AIC scores of m2 (intercepts for scenario) and m3 (intercepts for subjects) that random intercepts by subjects explain more variance than scenario. Using only random intercepts by subject as a second-level effect has a lower sigma value (32.04) and AIC score (2112.1) than by scenario.

2) Why is our single-level model bad?

The single level model ignores some important hierarchies in the data which we are well aware of exist. It disregards the fact that subjects might differ on baseline pitch, and that there may be differences across scenarios. Whilst it might explain some of the variance in the data, yes, it is to simple to be a complete answer.

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

```r
# Creating new df
politeness2 <- politeness %>%
  group_by(subject, gender) %>%
  summarize(mean_of_f0mn = mean(f0mn))
```

```
## `summarise()` has grouped output by 'subject'. You can override using the `.groups` argument.
```

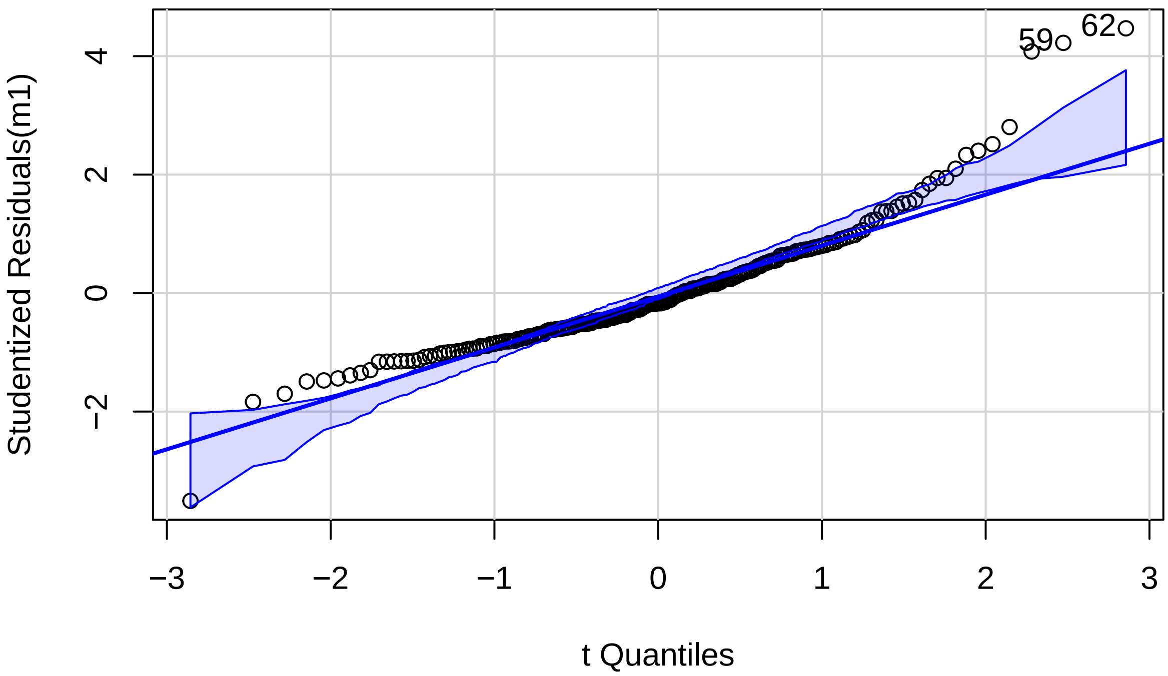
```
ii. build a single-level model that models _f0mn_ as dependent on _gender_ using this new dataset
```

```r
# Building new model
m5 <- lm(mean_of_f0mn ~ gender, data = politeness2)
summary(m5)
```

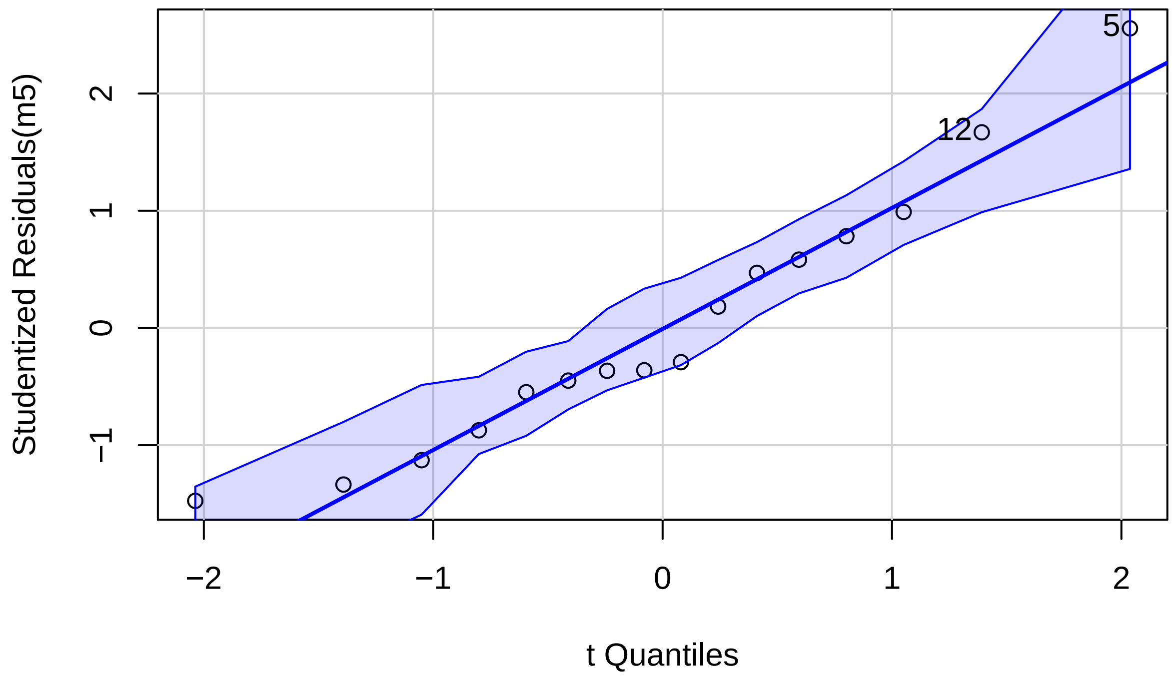
```
##
## Call:
## lm(formula = mean_of_f0mn ~ gender, data = politeness2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.606 -15.493  -8.212  15.702  52.859
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  246.370      8.635  28.530 8.34e-14 ***
## genderM     -115.092     13.055  -8.816 4.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.91 on 14 degrees of freedom
## Multiple R-squared:  0.8474, Adjusted R-squared:  0.8365
## F-statistic: 77.72 on 1 and 14 DF,  p-value: 4.346e-07
```

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

```r
par(car::qqPlot(m1), car::qqPlot(m5))
```



Studentized Residuals(m1)
4
2
0
−2
59
62
−3
−2
−1
0
1
2
3
t Quantiles




```
## [[1]]
## NULL
##
## [[2]]
## NULL
```

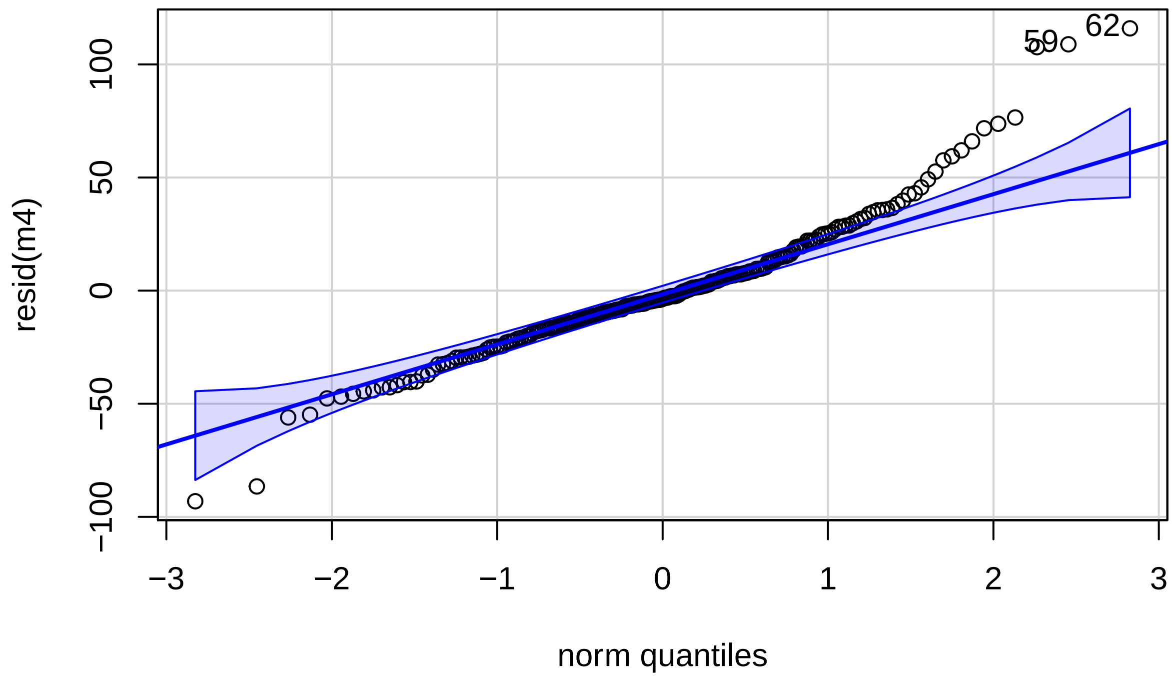
```
tibble(sigma(m1), sigma(m5))
```

```
## # A tibble: 1 x 2
##   ‘sigma(m1)‘ ‘sigma(m5)‘
##         <dbl>       <dbl>
## 1        39.5        25.9
```

The QQ-plot of the pooled f0 scores seems better than that of the individual scores. This can also be seen by the lower sigma value for the model with the pooled scores. Although it fulfills the assumptions of the model better, it reduces the resolution of the data by ignoring the difference of f0 between scenarios.

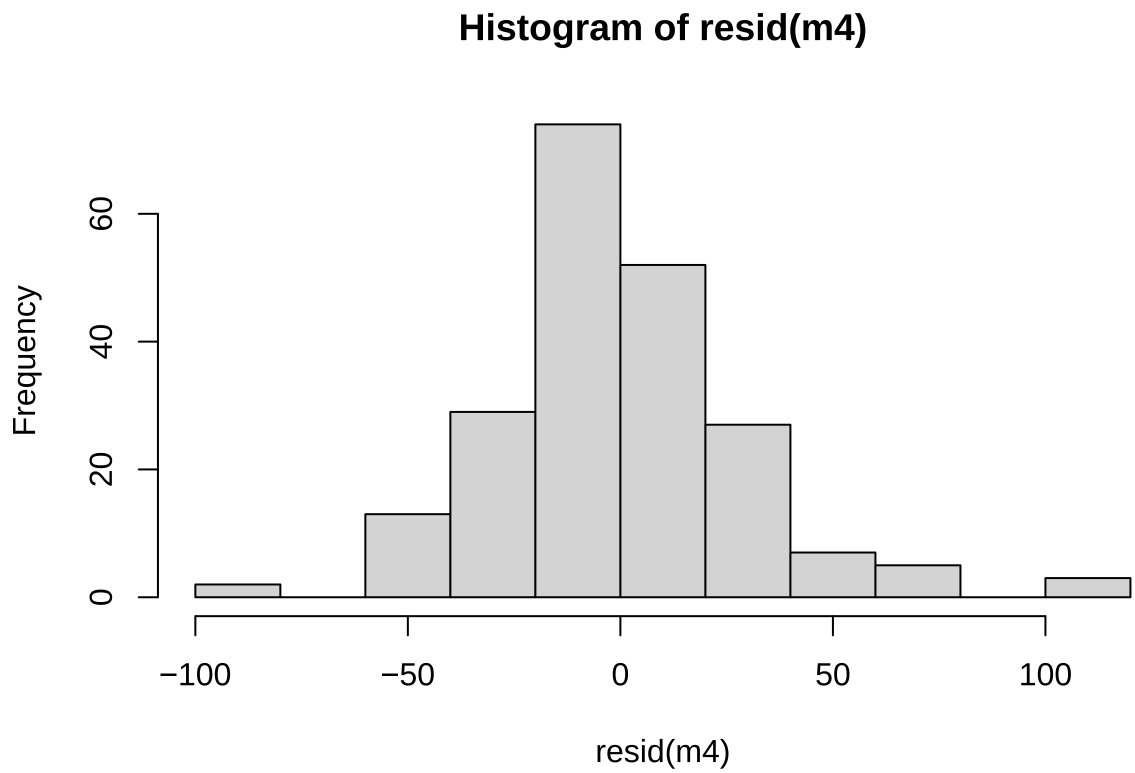
```
iv. Also make a quantile-quantile plot for the residuals of the  multilevel model with two intercepts.
```

```
# Plotting both qq plot and histogram of residuals
car::qqPlot(resid(m4))
```




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

```
hist(resid(m4))
```

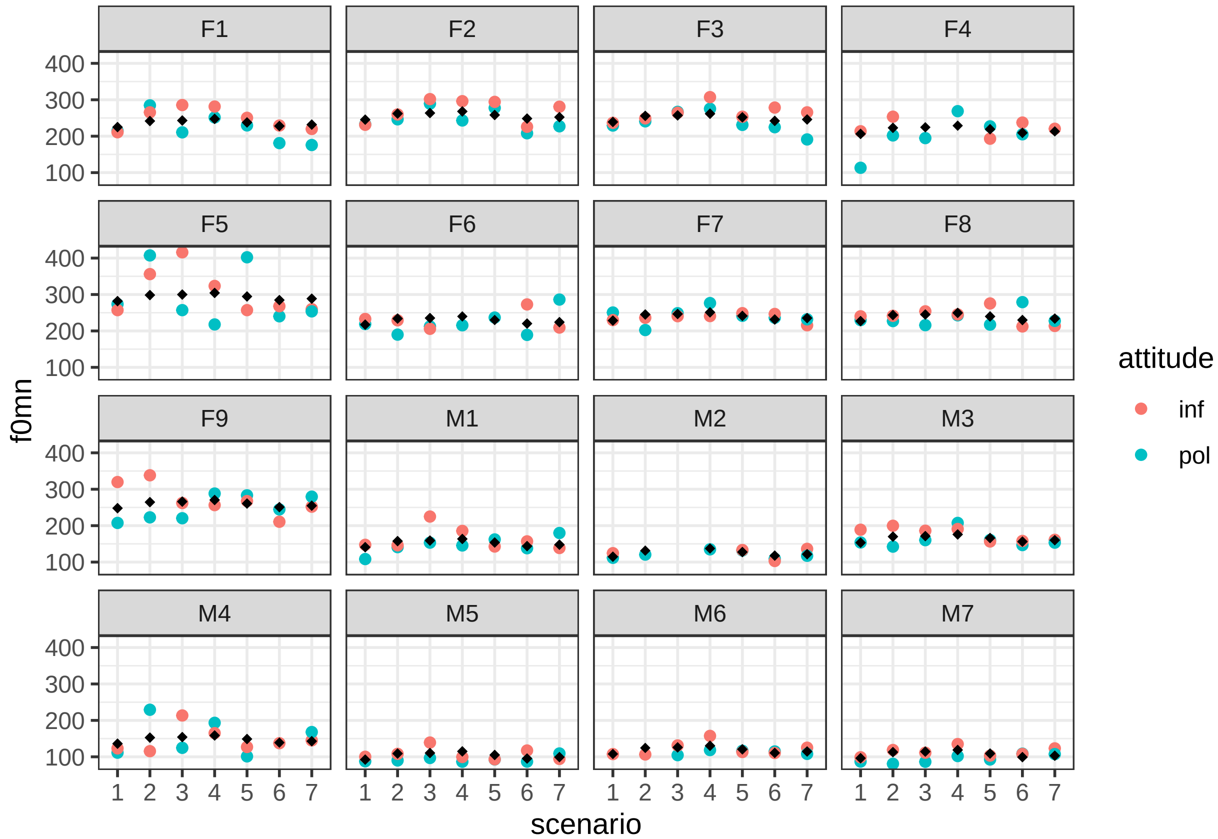


The residuals are a bit right-skewed, but it does look alright, perhaps a sleight violation of assumption.

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)

```
# Plotting
ggplot(data = politeness, aes(x = scenario, y = f0mn, color = attitude)) +
  geom_point() +
  geom_point(aes(x = scenario, y = fitted(m4)), color = "black", shape = 18)+
  facet_wrap(~subject)+
  theme_bw()
```



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 = F
summary(m6)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender + attitude + (1 | scenario) + (1 | subject)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##   2094.5   2114.6  -1041.2   2082.5      206
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8791 -0.5968 -0.0569  0.4260  3.9068
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 514.92   22.692
##  scenario (Intercept)  99.22    9.961
```

```
##  Residual             878.39   29.638
## Number of obs: 212, groups:  subject, 16; scenario, 7
##
## Fixed effects:
##              Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  254.408      9.117   21.800  27.904  < 2e-16 ***
## genderM     -115.447     12.161   16.000  -9.494 5.63e-08 ***
## attitudepol  -14.817      4.086  190.559  -3.626 0.000369 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.583
## attitudepol -0.231  0.006
```

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 = F
summary(m7)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ gender * attitude + (1 | scenario) + (1 | subject)
##    Data: politeness
##
##      AIC      BIC   logLik deviance df.resid
##   2096.0   2119.5  -1041.0   2082.0      205
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8460 -0.5893 -0.0685  0.3946  3.9518
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 514.09   22.674
##  scenario (Intercept)  99.08    9.954
##  Residual             876.46   29.605
## Number of obs: 212, groups:  subject, 16; scenario, 7
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)          255.632      9.289   23.556  27.521  < 2e-16 ***
## genderM             -118.251     12.841   19.922  -9.209 1.28e-08 ***
## attitudepol          -17.198      5.395  190.331  -3.188  0.00168 **
## genderM:attitudepol    5.563      8.241  190.388   0.675  0.50049
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM atttdp
## genderM     -0.605
## attitudepol -0.299  0.216
## gndrM:tttdp  0.195 -0.323 -0.654
```

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

The model describes that, although pitch decreases in the polite attitude compared to the informal attitude, Korean men's pitch seem (in our sample) to decrease less than women's pitch, though the interaction is not significant.

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.

```
# Running anova to quickly compare AIC values, making a df for sigmal values and one for SSR
anova(m4, m6, m7)
```

```
## Data: politeness
## Models:
## m4: f0mn ~ gender + (1 | scenario) + (1 | subject)
## m6: f0mn ~ gender + attitude + (1 | scenario) + (1 | subject)
## m7: f0mn ~ gender * attitude + (1 | scenario) + (1 | subject)
##    npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m4    5 2105.2 2122.0 -1047.6   2095.2
## m6    6 2094.5 2114.6 -1041.2   2082.5 12.6868  1  0.0003683 ***
## m7    7 2096.0 2119.5 -1041.0   2082.0  0.4551  1  0.4998998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
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
```

```
tibble(sum(resid(m4)^2), sum(resid(m6)^2), sum(resid(m7)^2))
```

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

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

Findings based on politeness dataset by Winter & Grawunder

Note on dataset

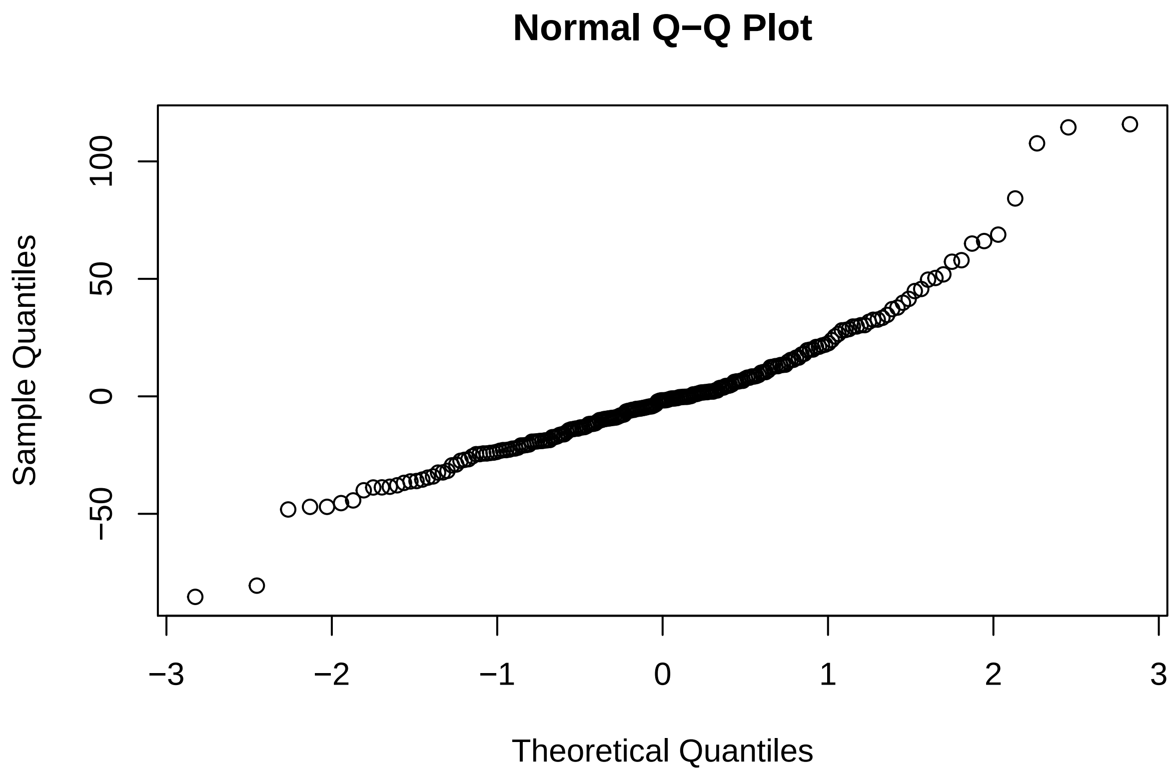
This dataset was collected to study the phonetic profile of Korean formal and informal speech registers (Winter & Grawunder, 2012). The dataset includes variables of: "Subject" - the anonymized participant ID; "Gender" - the gender of the participant, either F (female) or M (male); "Scenario" - categorized as 7 different scenarios of dialogue, labeled 1-7; "Attitude" - the attitude of the scenario, either formal or informal; "Total duration" - duration of the scenario; "f0mn" - the mean pitch of the participants voice in a scenario and "hiss_count" - how many hisses the subject utters during the scenario.

Building the model

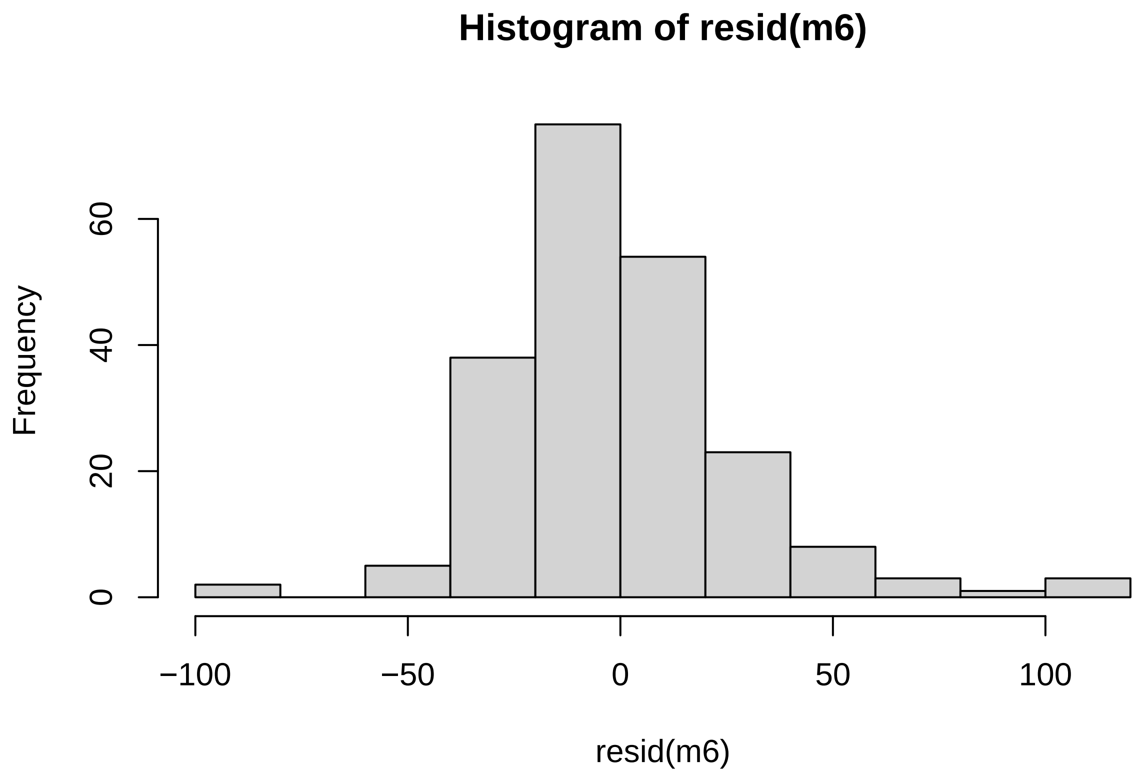
The variable of interest in the paper is mean pitch (f0mn), and I'd like to monitor the changes (if anyin) in this variable. Therefore, it is the outcome variable of my model. Since it is established that pitch is highly correlated to biological sex, I have included gender as a main (fixed) effect. As the original paper hypothesized about pitch differences in formal and informal settings, attitude should of course also be included as a main effect. Furthermore, it is likely that subject have varying baseline pitch frequencies, which is the case for adding a random intercepts per subject. The same goes for the variable scenario, which indicates the situation in which the dialogue exchange takes place. The input in R then becomes f0mn ~ gender + attitude + (1 | scenario) + (1 | subject). The main effects are both significant at $p < 0.05$, and the model also has the lowest AIC value of all tested models. There could be a case for adding the interaction between gender and attitude, as one might hypothesize whether there's a difference in degree of pitch change between men and women across attitude, though this was tested and yielded a weaker model with no significant effect of the interaction and a higher AIC value, and as also not particularly relevant to the study.

Checking assumptions of normality of residuals

```
qqnorm(resid(m6))
```




```
hist(resid(m6))
```



The residuals follow a somewhat normal distribution, with a slight right skew (positive skew), which slightly breaks assumptions.

Results from model

The predictor "gender" has a significant effect ($p < 0.05$) on mean pitch, showing a decrease from 254.41 Hz for women to 138.96 Hz for men. In comparing polite and informal speech, there is a 14.82 Hz reduction in mean pitch going from the informal to the polite condition ($p < 0.05$). We also see the random intercepts explaining a good amount of the variance in the dataset, especially random intercepts by subject (score of 514.9).