practical_exercise_2, Methods 3, 2021, autumn semester

Aleks

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
```

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

ALEKS:

Data comes from an experiment in which the voice pitch is investigated in the context of formal or informal interactions. 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 - Scenario: Different context of interaction (fx asking a professor for a favor vs asking a mate for a favor) - Attitude: 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`

```
## The first four is a factor, because they are categorial(more or less). And the 3 latter are factors a

politeness$subject = as.factor(politeness$subject)
politeness$gender = as.factor(politeness$gender)
politeness$scenario = as.factor(politeness$scenario)
politeness$attitude = as.factor(politeness$attitude)

## The latter three are already numeric and we can check it using is.numeric(df$column). I've done that

is.numeric(politeness$total_duration)
```

```
## [1] TRUE
```

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

```
df1 = filter(politeness, subject == "F1")

## As a factor.

model1 = lm(f0mn ~ scenario, df1)
summary(model1)
```

```
##
## Call:
## lm(formula = f0mn ~ scenario, data = df1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -37.50 -13.86   0.00  13.86  37.50
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   212.75      20.35  10.453  1.6e-05 ***
## scenario2      62.40      28.78   2.168   0.0668 .
## scenario3      35.35      28.78   1.228   0.2591
## scenario4      53.75      28.78   1.867   0.1041
## scenario5      27.30      28.78   0.948   0.3745
## scenario6      -7.55      28.78  -0.262   0.8006
## scenario7     -14.95      28.78  -0.519   0.6195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.78 on 7 degrees of freedom
## Multiple R-squared:  0.6576, Adjusted R-squared:  0.364
## F-statistic:  2.24 on 6 and 7 DF,  p-value: 0.1576
```

```r
## Now as integer.
df1$scenario = as.integer(df1$scenario)
model2 = lm(f0mn ~ scenario, df1)
summary(model2)
```

```
##
## Call:
## lm(formula = f0mn ~ scenario, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.836 -36.807   6.686  20.918  46.421
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  262.621     20.616  12.738 2.48e-08 ***
## scenario      -6.886      4.610  -1.494    0.161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.5 on 12 degrees of freedom
## Multiple R-squared:  0.1568, Adjusted R-squared:  0.0865
## F-statistic: 2.231 on 1 and 12 DF,  p-value: 0.1611
```

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
model.matrix(model1)
```

```
##    (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"
```

```r
model.matrix(model2)
```

```
##    (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
```

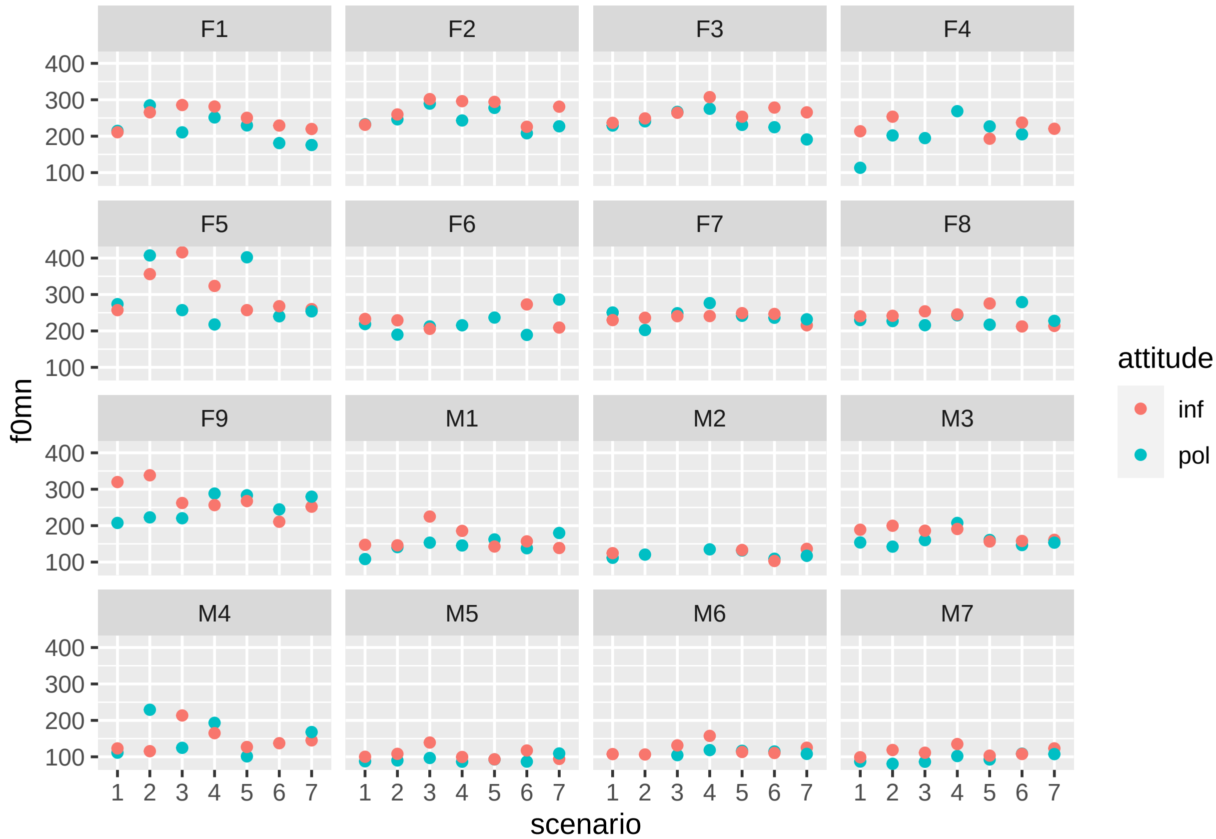
ii. Which coding of *scenario*, as a factor or not, is more fitting?

ALEKS: You want to treat all scenarios separately, so factors are more fitting than integers.

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

```
## Warning: Removed 12 rows containing missing values (geom_point).
```



i. Describe the differences between subjects

ALEKS: Well, as we can see the males have generally lowers pitches (as expected, I guess) and there doesn't really seem to be that big of a difference between attitudes, right?

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*

```
model1 = lm(f0mn ~ gender, politeness)
model2 = lmer(f0mn ~ gender + (1|scenario) , politeness, REML = FALSE)
model3 = lmer(f0mn ~ gender + (1|subject) , politeness, REML = FALSE)
model4 = lmer(f0mn ~ gender + (1|scenario) + (1|subject) , politeness, REML = FALSE)
```

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

```
#Residual standard deviation
which.min(tibble(sigma(model1), sigma(model2), sigma(model3), sigma(model4)))
```

```
## sigma(model4)
##             4
```

```
head(tibble(sigma(model1), sigma(model2), sigma(model3), sigma(model4)))
```

```
## # A tibble: 1 x 4
##   `sigma(model1)` `sigma(model2)` `sigma(model3)` `sigma(model4)`
##             <dbl>           <dbl>           <dbl>           <dbl>
## 1            39.5            38.4            32.0            30.7
```

```
#AIC
head(tibble(AIC(model1), AIC(model2), AIC(model3), AIC(model4)))
```

```
## # A tibble: 1 x 4
##   `AIC(model1)` `AIC(model2)` `AIC(model3)` `AIC(model4)`
##           <dbl>         <dbl>         <dbl>         <dbl>
## 1         2164.         2162.         2112.         2105.
```

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

ALEKS: Unique intercept per *subject* is the second-level effect that explains the most variance, as seen by both residual standard error and AIC comparisons.

2) Why is our single-level model bad?

ALEKS: Well, it's because it does not account for random effects(such as different pitch baseline per participant or scenario) - which, in the context of this experiment, make a lot of sense to account for.

i. create a new data frame that has three variables, *subject*, *gender* and *f0mn*, where *f0mn* is the average of all responses of each subject, i.e. averaging across *attitude* and_scenario_

```
df2 = politeness %>%
  group_by(subject, gender) %>%
  summarize(mean_f0mn = mean(f0mn))
```

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

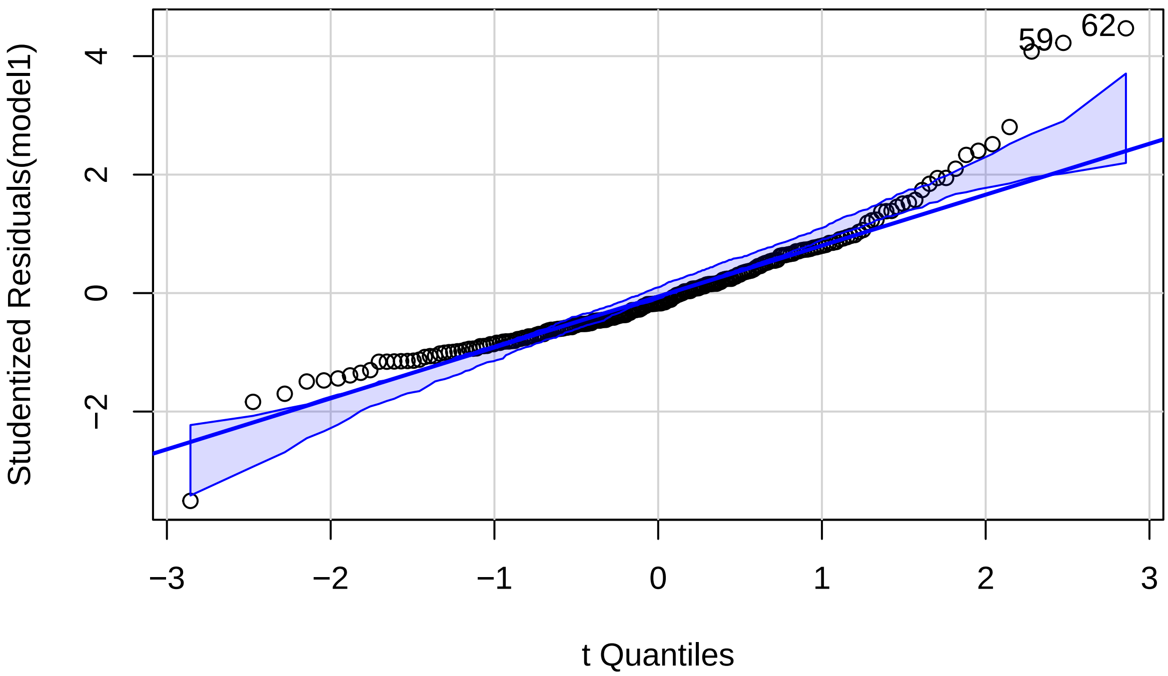
ii. build a single-level model that models *f0mn* as dependent on *gender* using this new dataset

```r
model5 = lm(mean_f0mn ~ gender, df2)
summary(model5)
```

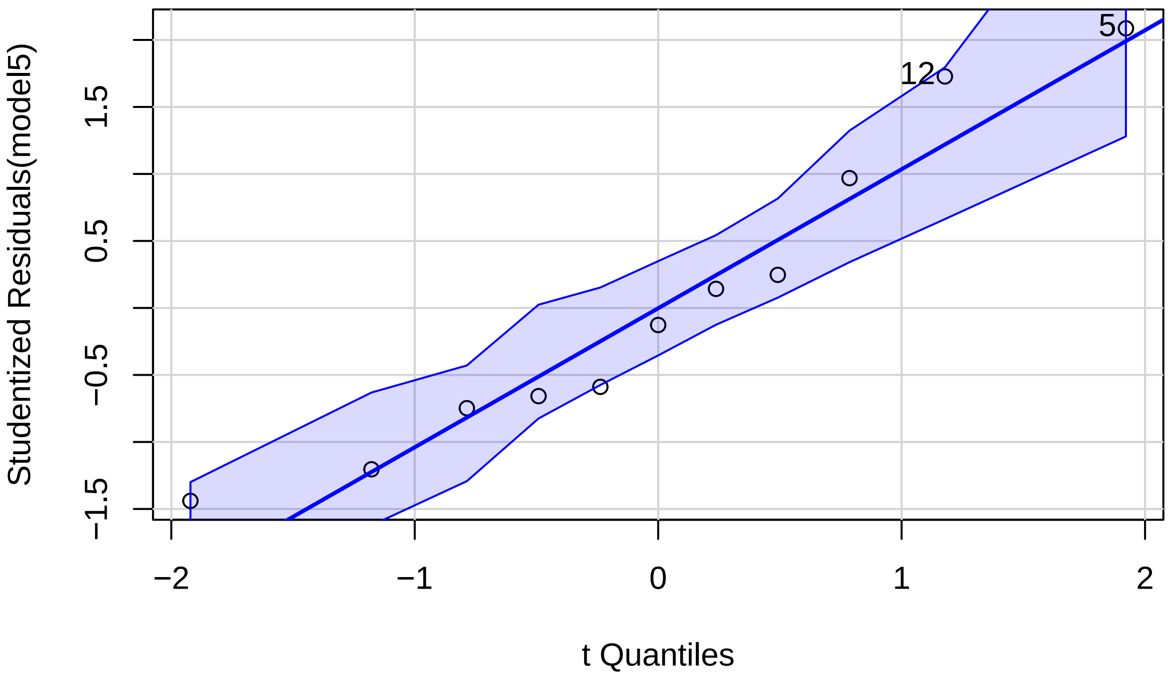
```
##
## Call:
## lm(formula = mean_f0mn ~ gender, data = df2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.050 -18.212  -3.394  14.732  44.842
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   254.39      10.28  24.749 1.38e-09 ***
## genderM      -122.24      17.05  -7.171 5.24e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.19 on 9 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.8511, Adjusted R-squared:  0.8345
## F-statistic: 51.43 on 1 and 9 DF,  p-value: 5.243e-05
```

iii. make Quantile-Quantile plots, comparing theoretical quantiles to the sample quantiles) using `qqnorm` and `qqline` for the new single-level model and compare it to the old single-level model (from 1).i). Which model's residuals (ϵ) fulfil the assumptions of the General Linear Model better?)

```r
par(car::qqPlot(model1), car::qqPlot(model5))
```



Studentized Residuals(model1)
t Quantiles
59
62
4
2
0
−2
−3
−2
−1
0
1
2
3




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

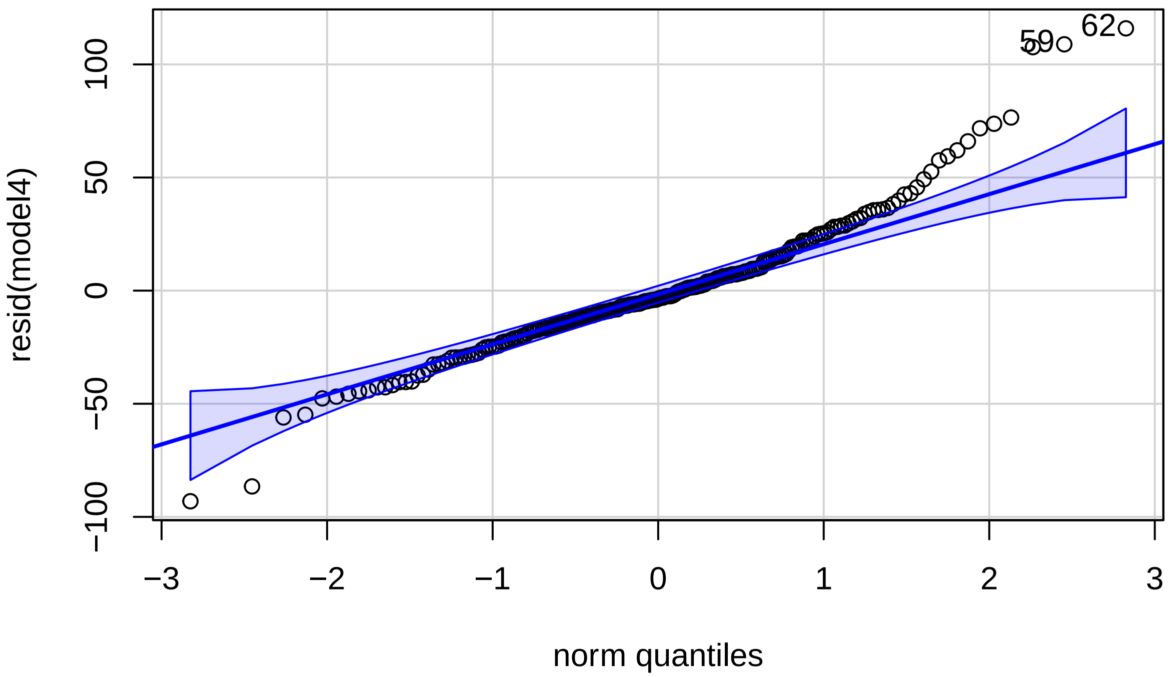
```
tibble(sigma(model1), sigma(model5))
```

```
## # A tibble: 1 x 2
##   `sigma(model1)` `sigma(model5)`
##             <dbl>           <dbl>
## 1            39.5            27.2
```

ALEKS:The qqplot for the model from pooled *f0mn* scores seems better, and the sigma value for this model is also lower.

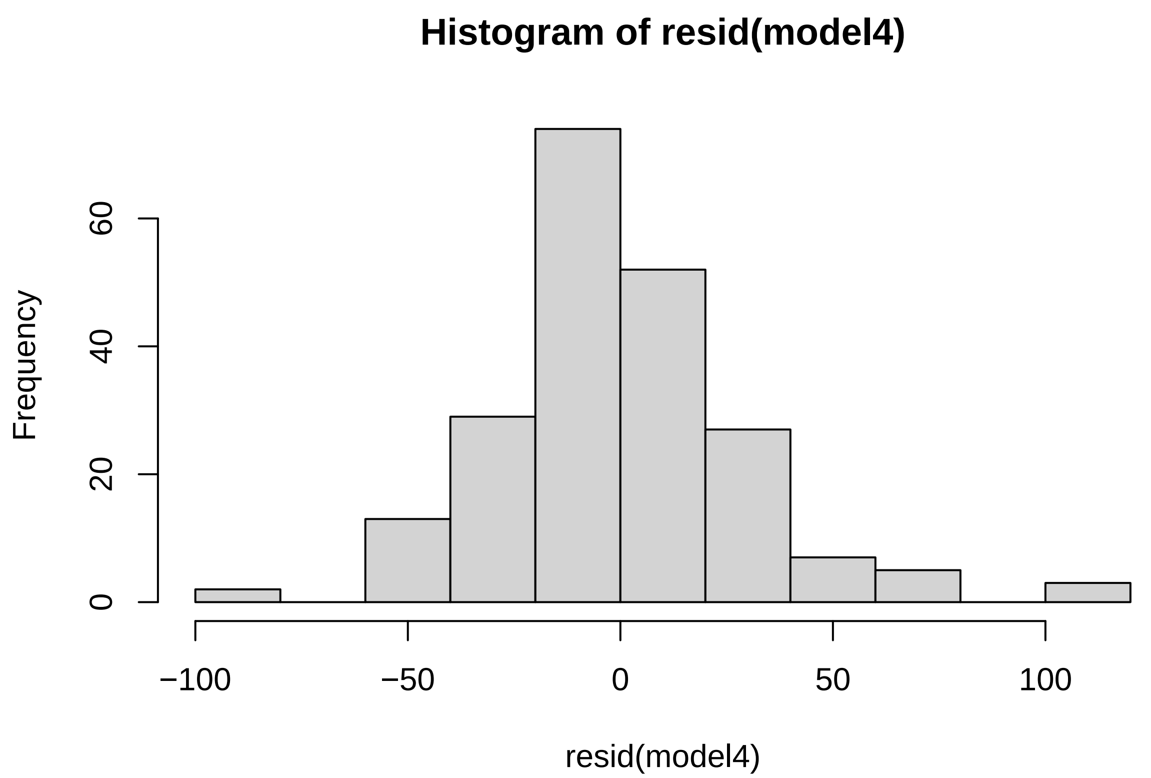
iv. Also make a quantile-quantile plot for the residuals of the multilevel model with two intercepts. Does it look alright?

```
car::qqPlot(resid(model4))
```




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

```
hist(resid(model4))
```



ALEKS: The residuals are slightly right-skewed, but I guess that is within the boundaries of acceptance.

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)

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*

```
model4 = lmer(f0mn ~ gender + (1|scenario) + (1|subject) , politeness, REML = FALSE)
model6 = lmer(f0mn ~ gender + attitude + (1|scenario) + (1|subject), politeness, REML = FALSE)
```

ii. make a separate model that besides the main effects of *attitude* and *gender* also include their interaction

```
model7 = lmer(f0mn ~ gender + attitude + gender*attitude + (1|scenario) + (1|subject), politeness, REML
summary(model2)
```

```
## 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. describe what the interaction term in the model says about Korean men's pitch when they are polite relative to Korean women's pitch when they are polite (you don't have to judge whether it is interesting)

ALEKS: For both genders the pitch decreases in an informal scenario, but for men it decreases less than for women. Not significant though.

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.

```r
#Residual variance
tibble(sum(resid(model4)^2), sum(resid(model6)^2), sum(resid(model7)^2))
```

```
## # A tibble: 1 x 3
##   `sum(resid(model4)^2)` `sum(resid(model6)^2)` `sum(resid(model7)^2)`
##                    <dbl>                  <dbl>                  <dbl>
## 1                181913.                169681.                169306.
```

```r
#Residual standard deviation
head(tibble(sigma(model4), sigma(model6), sigma(model7)))
```

```
## # A tibble: 1 x 3
##   `sigma(model4)` `sigma(model6)` `sigma(model7)`
##             <dbl>           <dbl>           <dbl>
## 1            30.7            29.6            29.6
```

```r
#AIC
head(tibble(AIC(model4), AIC(model6), AIC(model7)))
```

```
## # A tibble: 1 x 3
##   `AIC(model4)` `AIC(model6)` `AIC(model7)`
##           <dbl>         <dbl>         <dbl>
## 1         2105.         2094.         2096.
```

ALEKS: Interesting, the residual standard deviation is smallest for Model 7(that includes the interaction between gender and attitude), however, the AIC of this model is slightly higher than the one where interaction is not modelled. Could this be the case for overfitting? Conceptually, it could be the case that speaking formally (vs informally) has different effect on men (pitch goes up) than on women (pitch goes down). As such, I think this model is the best one, even if its AIC is slightly higher.

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:

ALEKS: I hereby choose model6 as best model, characterized by lowest AIC score and near-lowest residual variance and residual standard error. Whilst model7 is slightly lower on the latter two characteristics, I choose to use model6 as it is less complex and as such, less prone to overfitting.

```r
summary(model6)
```

```
## 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
```

i. describe what the dataset consists of

ALEKS:

Data comes from an experiment in which the voice pitch is investigated in the context of formal or informal interactions. 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 - Scenario: Different context of interaction (fx asking a professor for a favor vs asking a mate for a favor) - Attitude: 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

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

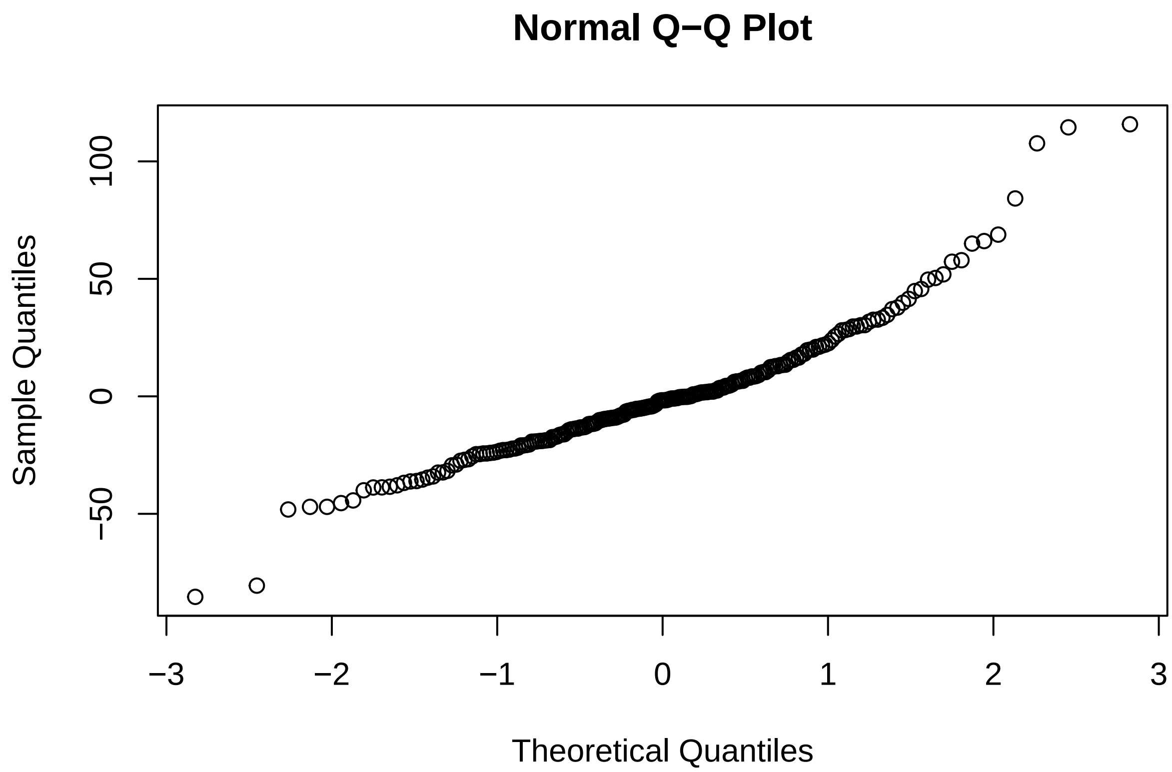
Both gender and attitude modulate the pitch.

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

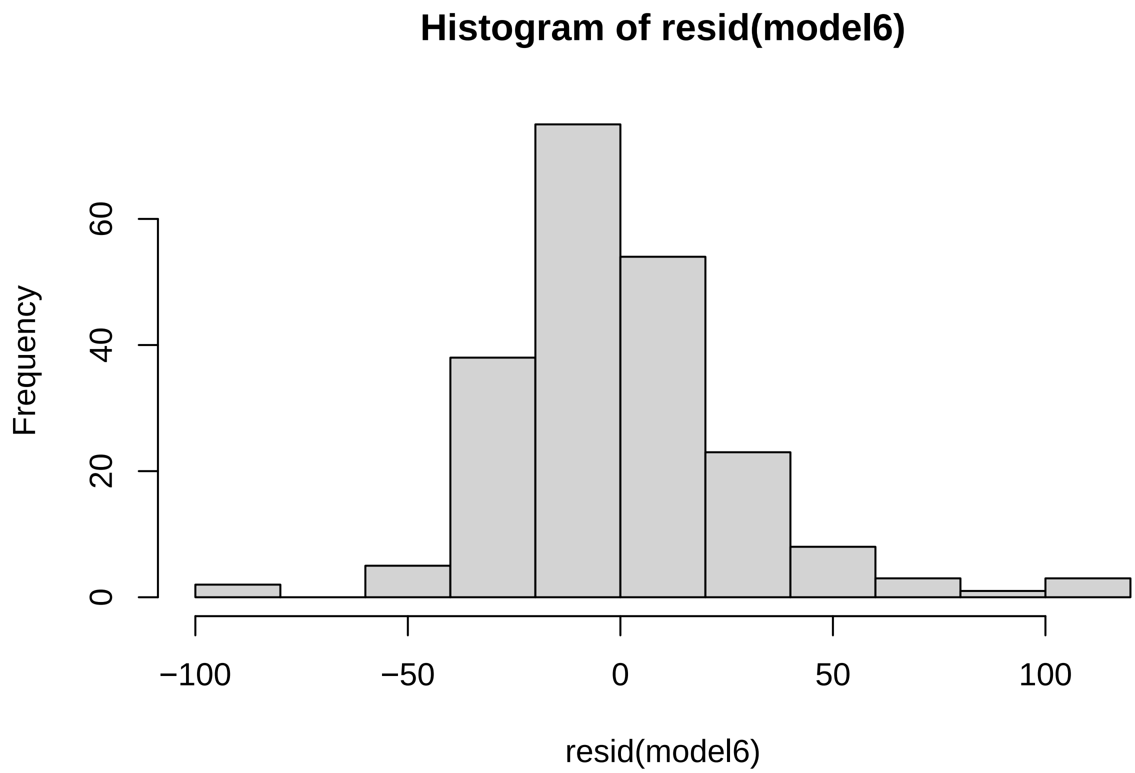
ALEKS: You can expect a different pitch baseline for each subject(people differ) and you can expect different scenarios to have a different pitch baseline as well(an "excusing yourself" scenario might have a different pitch baseline when comapred to "asking for a favor" scenario).

ALEKS: WHAT?
v. include a Quantile-Quantile plot of your chosen model

```
qqnorm(resid(model6))
```




```
hist(resid(model6))
```



Results from model

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. A reduction of 14.82 Hz in mean pitch is observed going from the informal to the formal scenario ($p < 0.05$). The random effects explain a good amount of the variance in the dataset, especially random intercepts by subject (score of 514.9).