practical_exercise_2, Methods 3, 2021, autumn semester

[FILL IN YOUR NAME]

[FILL IN THE DATE]

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.

```
pacman::p_load(rstanarm, tidyverse, lmerTest, lme4)
pacman::p_load(MuMIn)
```

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

```
summary(politeness)
```

```
##    subject             gender            scenario   attitude
##  Length:224         Length:224         Min.   :1   Length:224
##  Class :character   Class :character   1st Qu.:2   Class :character
##  Mode  :character   Mode  :character   Median :4   Mode  :character
##                                        Mean   :4
##                                        3rd Qu.:6
##                                        Max.   :7
##
##  total_duration         f0mn         hiss_count
##  Min.   :  0.988   Min.   : 80.8   Min.   :0.0000
##  1st Qu.:  6.140   1st Qu.:134.5   1st Qu.:0.0000
##  Median : 12.640   Median :211.6   Median :0.0000
##  Mean   : 24.176   Mean   :197.9   Mean   :0.4509
##  3rd Qu.: 39.373   3rd Qu.:247.1   3rd Qu.:1.0000
##  Max.   :101.375   Max.   :415.8   Max.   :5.0000
##                    NA's   :12
```

```
#Change variables to appropiate factors
politeness <- politeness %>%
  mutate(scenario = as.factor(scenario)) %>%
  mutate(gender = as.factor(gender)) %>%
  mutate(attitude = as.factor(attitude))
head(politeness)
```

```
##   subject gender scenario attitude total_duration  f0mn hiss_count
## 1      F1      F        1      pol         18.392 214.6          2
## 2      F1      F        1      inf         13.551 210.9          0
## 3      F1      F        2      pol          5.217 284.7          0
## 4      F1      F        2      inf          4.247 265.6          0
## 5      F1      F        3      pol          6.791 210.6          0
## 6      F1      F        3      inf          4.126 285.6          0
```

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

```
politeness_F1 <- politeness %>%
  filter(subject == "F1")
m1 <- lm(f0mn ~ scenario, data = politeness_F1, refresh = 0)
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'refresh' will be disregarded
```

```
m2 <- lm(f0mn ~ as.integer(scenario), data = politeness_F1, refresh = 0)
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'refresh' will be disregarded
```

```
i. Include the model matrices, $X$ from the General Linear Model, for these two models in your report ar
```

Scenario Encoded as Factor

```
#Design Matrix
model.matrix(m1)
```

```
##    (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"
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = f0mn ~ scenario, data = politeness_F1, refresh = 0)
##
## 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
```

The design matrix is a [14x7] matrix, so we will get the following β_{0-6}. This is also shown by the summary of a our linear regression model. A simple regression f0mn ~ scenario was conducted. Scenario seemed to account for 36.4% of the variance in f0mn following adjusted R^2. F(1,6) = 2.24, p >0.5) all beta values were insignificant. We only have 14 observations spread out over 7 different levels. So the high p-value is most likely due to sample-size. A further power-analysis could show the required sample size required.

Scenario Encoded as Int

```r
#Design Matrix
model.matrix(m2)
```

```
##    (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
```

```r
summary(m2)
```

```
##
## Call:
## lm(formula = f0mn ~ as.integer(scenario), data = politeness_F1,
##     refresh = 0)
##
## 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 ***
## as.integer(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
```

Now that scenario is encoded as an integer the design matrix will be a [14x2] matrix. Our model will therefore only give us β_{0-1} and not a β for each level of scenario as done in the previous model. This model assumes that there is a constant increment of f0mn following a "increase" in scenario (if you can even talk about a unit increase of scenario). This would only make sense if scenarios were ordered as getting harder and harder. The model is again f0mn ~ scenario F(1,12) = 2.231, p>0.5) with an adjusted $R^2 = 0.0865$ showing an explained variance of 8.65% (β_1 = -6.886, SE = 4.6, t = -1.5, p>0.16.) Again such a small sample size might be tricky to work with.

```
ii. Which coding of _scenario_, as a factor or not, is more fitting?
```

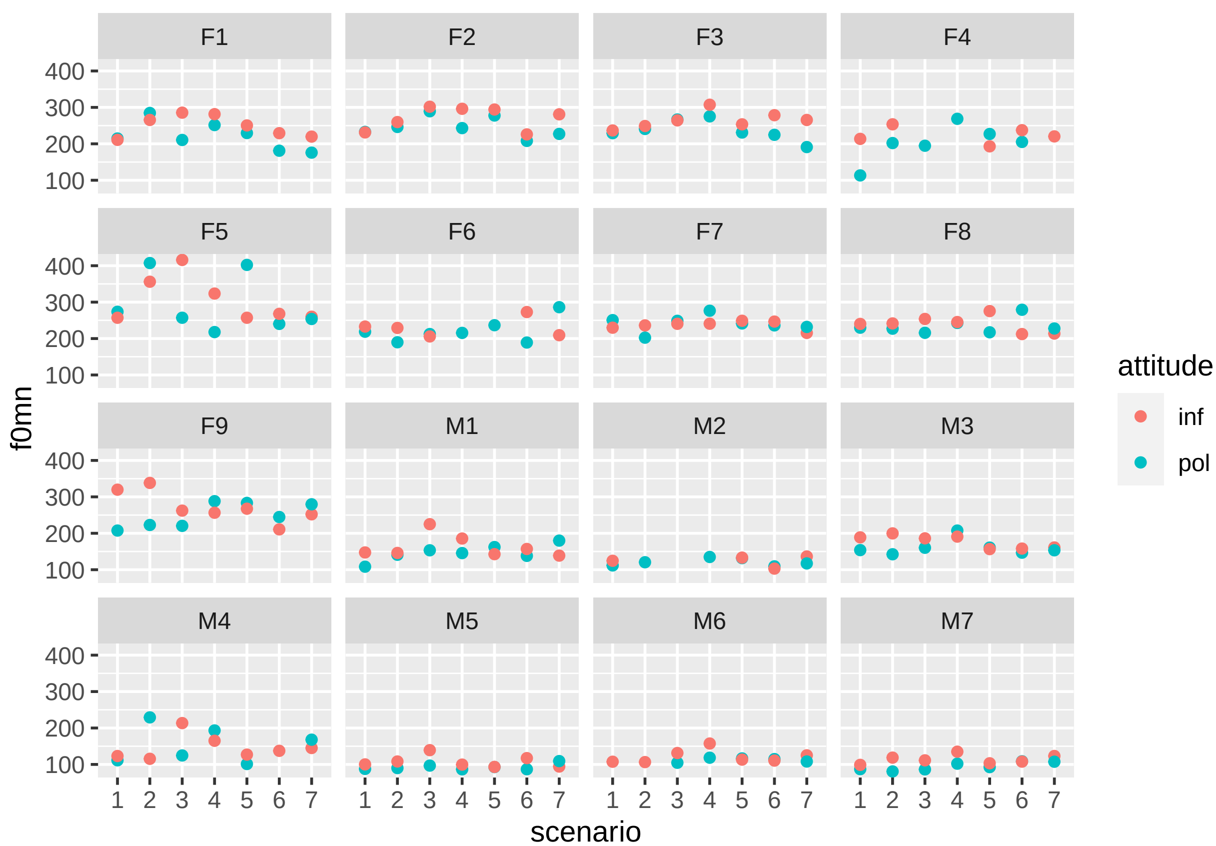
I would argue that scenario treated as a factor makes more sense. As mentioned a linear relationship between scenario number and f0mn does not make sense.

Scenario's effect on f0mn in such a scenario as, Scenario[1] < Scenario[2] > Scenario[3]. Would not be possible to model having scenario as an integer.

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

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



There seem to be a lower baseline/intercept given that you're a male. Attitude doesn't seem to have an large effect on f0mn. So an idea could be to add Gender as a fixed effect and subject as a random intercept as there is also individual variance within the gender category.

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(f0mn ~ scenario + gender + (1|subject), data=politeness)
```

1) Build four models and do some comparisons
    i. a single level model that models *f0mn* as dependent on *gender*

```
m3.1 <- lm(f0mn ~ gender, data = politeness)
```

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

```
m3.2 <- lmer(f0mn ~ gender + (1|scenario), data = politeness, REML = F)
```

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

```
m3.3 <- lmer(f0mn ~ gender + (1|subject), data = politeness, REML = F)
```

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

```
m3.4 <- lmer(f0mn ~ gender + (1|scenario) + (1|subject), data = politeness, REML = F)
```

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

```
#Calculate Residual Standard Deivation
c(S_res1 = sqrt(sum(resid(m3.1)^2)/(nrow(politeness)-2)),
S_res2 = sqrt(sum(resid(m3.2)^2)/(nrow(politeness)-2)),
S_res3 = sqrt(sum(resid(m3.3)^2)/(nrow(politeness)-2)),
S_res4 = sqrt(sum(resid(m3.4)^2)/(nrow(politeness)-2)))
```

```
##   S_res1   S_res2   S_res3   S_res4
## 38.38130 37.10991 30.27006 28.62566
```

```
#Compare AIC
AIC(m3.1,m3.2,m3.3,m3.4)
```

```
##      df      AIC
## m3.1  3 2163.971
## m3.2  4 2162.257
## m3.3  4 2112.048
## m3.4  5 2105.176
```

m3.4 is the model with the lowest residual standard deviation and also performs the best following the AIC.

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

```
#Anova cannot compare multi-level and single-level models. :(
anova(m3.2,m3.3,m3.4)
```

```
## Data: politeness
## Models:
## m3.2: f0mn ~ gender + (1 | scenario)
## m3.3: f0mn ~ gender + (1 | subject)
## m3.4: f0mn ~ gender + (1 | scenario) + (1 | subject)
##      npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m3.2    4 2162.3 2175.7 -1077.1   2154.3
## m3.3    4 2112.1 2125.5 -1052.0   2104.1 50.2095  0
## m3.4    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
```

```
#Look for varaince explained
MuMIn::r.squaredGLMM(m3.2)
```

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
##            R2m       R2c
## [1,] 0.6817304 0.6965456
```

```
MuMIn::r.squaredGLMM(m3.3)
```

```
##            R2m       R2c
## [1,] 0.6798832 0.7862932
```

```
MuMIn::r.squaredGLMM(m3.4)
```

```
##            R2m       R2c
## [1,] 0.6787423 0.8045921
```

M3.2 (f0mn ~ gender + (1|scenario)) showed the best variance explained purely by fixed effects. But m3.4 (f0mn ~ gender + (1|scenario) + (1|subject)) showed most explained variance with 80% of the variance being accounted for from both fixed and random effects.

We can also conclude that adding subject as random intercept rather than scenario explains more of the variance but also has more shared variance with our fixed effect gender.

2) Why is our single-level model bad?
    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_

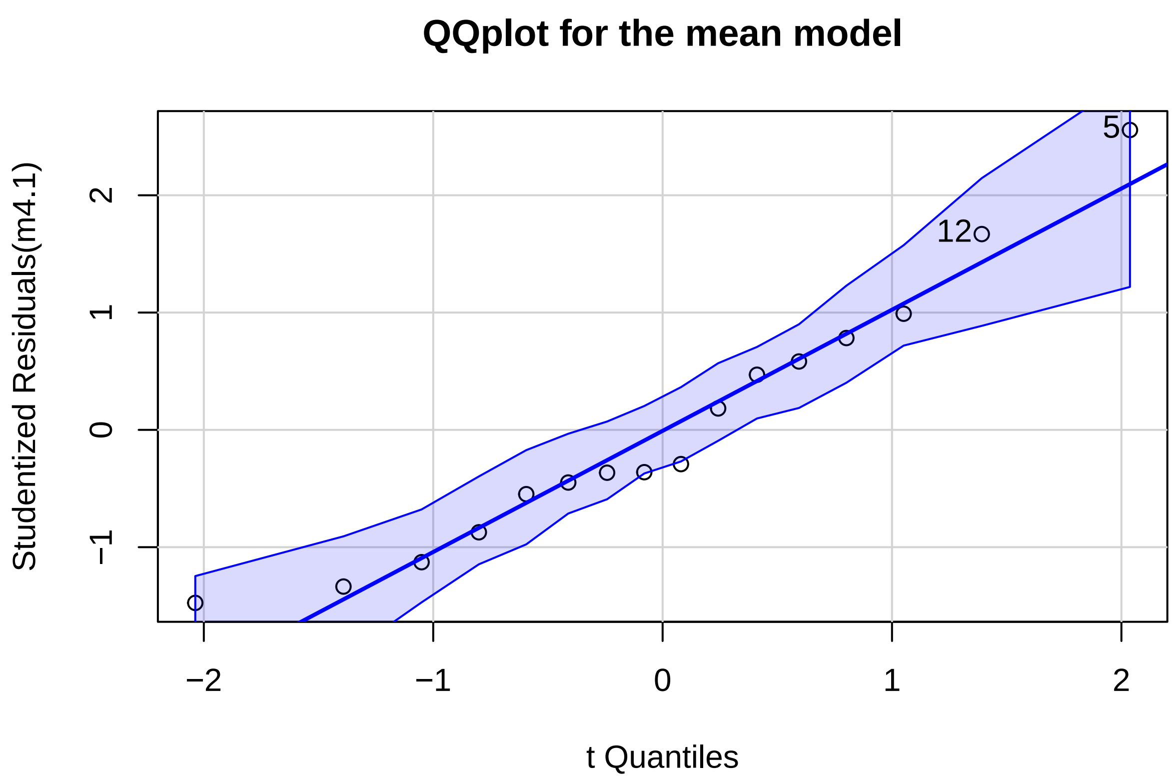
```
politeness_sel <- politeness %>%
  filter(!is.na(f0mn)) %>%
  group_by(subject) %>%
  summarise(f0mn = mean(f0mn))

politeness_sel <- politeness_sel %>%
  mutate(gender = if_else(grepl("F", politeness_sel$subject, ignore.case = T),"F","M")) %>%
  mutate(gender = as.factor(gender))
```

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

```
m4.1 <- lm(f0mn ~ gender, data = politeness_sel)
```

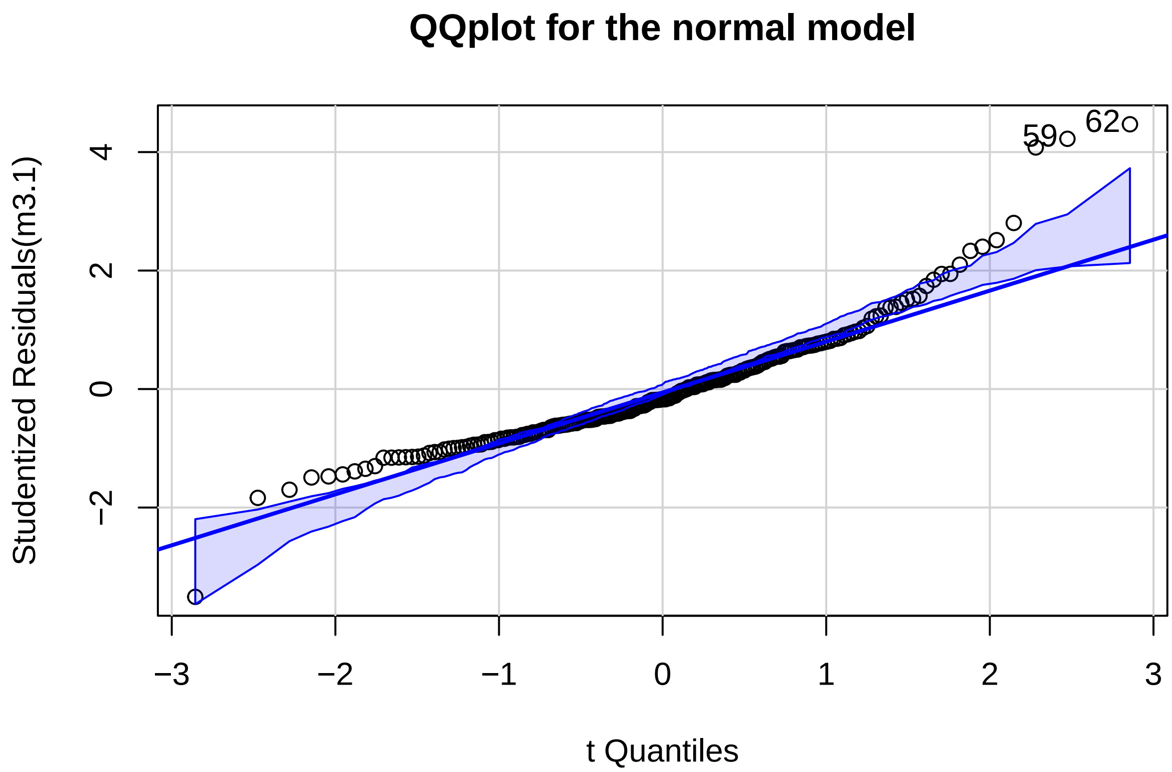
```
iii. make Quantile-Quantile plots, comparing theoretical quantiles to the sample quantiles) using ‘qqnor
```

```
#qqPlot
car::qqPlot(m4.1, main = "QQplot for the mean model")
```




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

```
car::qqPlot(m3.1, main = "QQplot for the normal model")
```



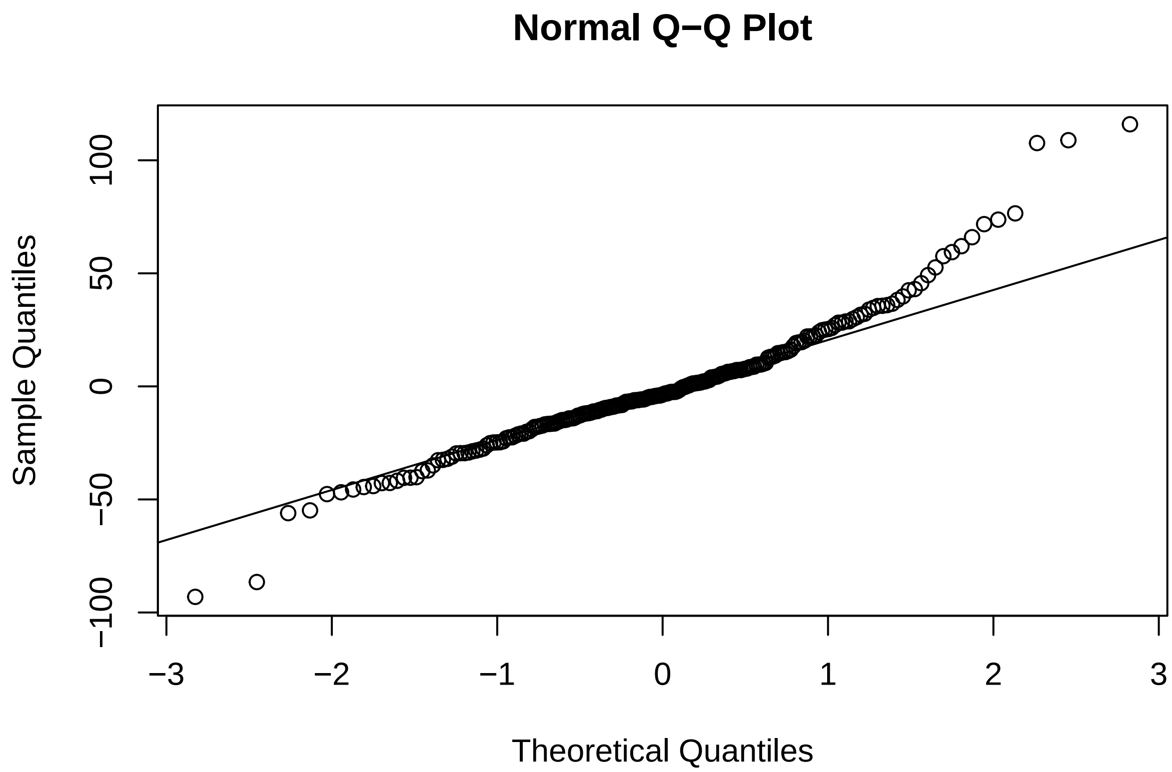

```
## [1] 59 62
```

Both models residual distribution seems to be slighly heavy tailed. This could be solved with a simple yuen trim of outliers. The first model 3.1 holds the assumptions for the GLM better than the new model where f0mn is averaged.

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

3) Plotting the two-intercepts model

```
#car::qqPlot doesn't like mixed effect models so we do it like this.
qqnorm(resid(m3.4))
qqline(resid(m3.4))
```




```
i. Create a plot for each subject, (similar to part 3 in Exercise 1), this time also indicating the fit
```

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*

```
m5.1 <- lmer(f0mn ~ attitude + gender + (1|subject) + (1|scenario), data = politeness, REML = F)
```

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

```
m5.2 <- lmer(f0mn ~ attitude*gender + (1|subject)+ (1|scenario), data = politeness, REML = F)
```

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

```
summary(m5.2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ attitude * gender + (1 | subject) + (1 | scenario)
##    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 ***
## attitudepol          -17.198      5.395  190.331  -3.188  0.00168 **
## genderM             -118.251     12.841   19.922  -9.209 1.28e-08 ***
## attitudepol:genderM    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) atttdp gendrM
## attitudepol -0.299
## genderM     -0.605  0.216
## atttdpl:gnM  0.195 -0.654 -0.323
```

```
MuMIn::r.squaredGLMM(m5.2)
```

```
##            R2m       R2c
## [1,] 0.6904904 0.8178935
```

```
levels(politeness$gender)
```

```
## [1] "F" "M"
```

```
levels(politeness$attitude)
```

```
## [1] "inf" "pol"
```

The model f0mn ~ attitude:gender + (1|subject)+ (1|scenario)has an R^2c 0.81 both attitude and gender showed a significant effect on f0mn (β_1(attitude_pol) = -17.2, SE = 5.4, p>0.05) and (β_2(genderM) = -119, SE = 12.8, p>0.05). Being polite and male lowers your frequency. Being both Male and Polite has an interaction effect of (β_3 = 5.5, SE = 8.24, p<0.05). Hereby concluding that there is a small positive insignificant interaction effect of being male and polite. The SE being proportional large compared to the effect size makes it very difficult to say anything meaningfull.

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
#reidual variance
c(RS_5.2 = summary(m5.2)$sigma^2,
RS_5.1 = summary(m5.1)$sigma^2,
RS_3.4 = summary(m3.4)$sigma^2)
```

```
##   RS_5.2   RS_5.1   RS_3.4
## 876.4591 878.3941 940.2532
```

```r
#residual standard deviation
c(S_res5.2 = sqrt(sum(resid(m5.2)^2)/(nrow(politeness)-2)),
S_res5.1 = sqrt(sum(resid(m5.2)^2)/(nrow(politeness)-2)),
S_res3.4 = sqrt(sum(resid(m3.4)^2)/(nrow(politeness)-2)))
```

```
## S_res5.2 S_res5.1 S_res3.4
## 27.61590 27.61590 28.62566
```

```r
anova(m3.4, m5.1, m5.2)
```

```
## Data: politeness
## Models:
## m3.4: f0mn ~ gender + (1 | scenario) + (1 | subject)
## m5.1: f0mn ~ attitude + gender + (1 | subject) + (1 | scenario)
## m5.2: f0mn ~ attitude * gender + (1 | subject) + (1 | scenario)
##      npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m3.4    5 2105.2 2122.0 -1047.6   2095.2
## m5.1    6 2094.5 2114.6 -1041.2   2082.5 12.6868  1  0.0003683 ***
## m5.2    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
```

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

My answer to all of the above

I have selected the model 5.1 (f0mn ~ gender + attitude + (1|subject) + (1|scenario)) My decision is based primary based on AIC R_res and RS. But theoretically it also makes sense to include both random intercepts due to the study being repeated measure and some variance being random/unsystematic. attitude furthermore seems like an important addition to the model as specific attitudes are correlated with pitch frequency (See imaginiry study). However the interaction between gender:attitude doesn't add anything explaining to the model.

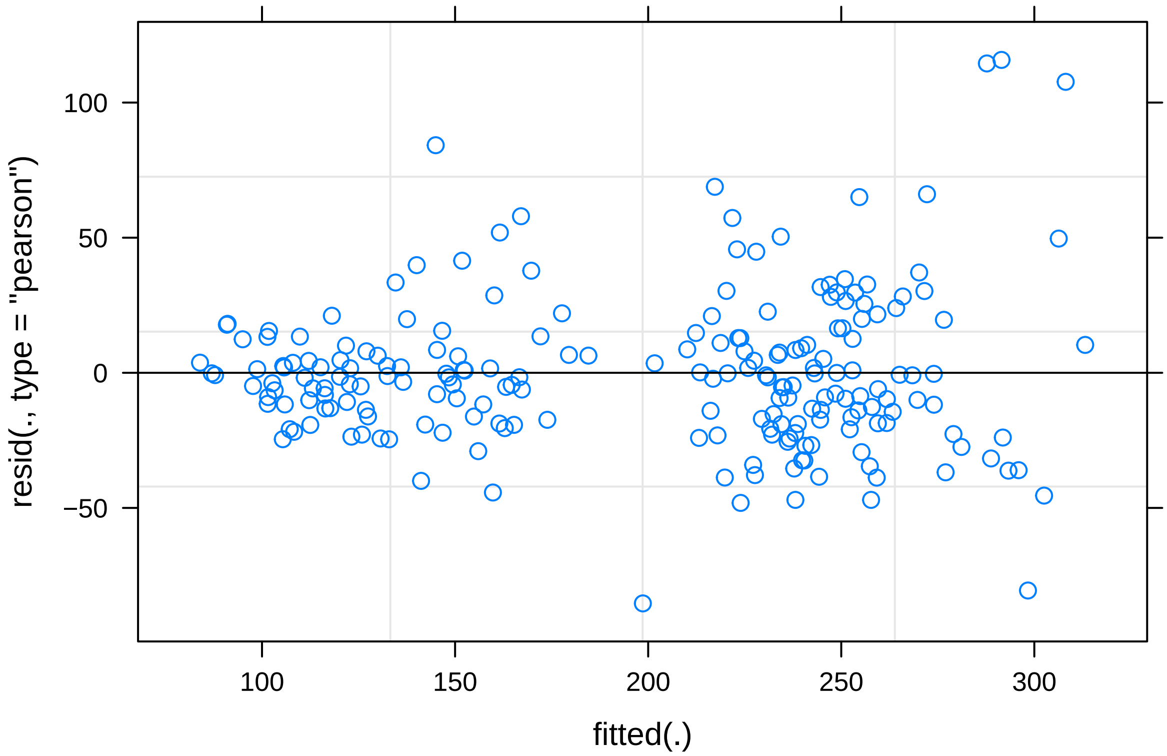
```
summary(m5.1)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: f0mn ~ attitude + gender + (1 | subject) + (1 | scenario)
##    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 ***
## attitudepol  -14.817      4.086  190.559  -3.626 0.000369 ***
## genderM     -115.447     12.161   16.000  -9.494 5.63e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) atttdp
## attitudepol -0.231
## genderM     -0.583  0.006
```

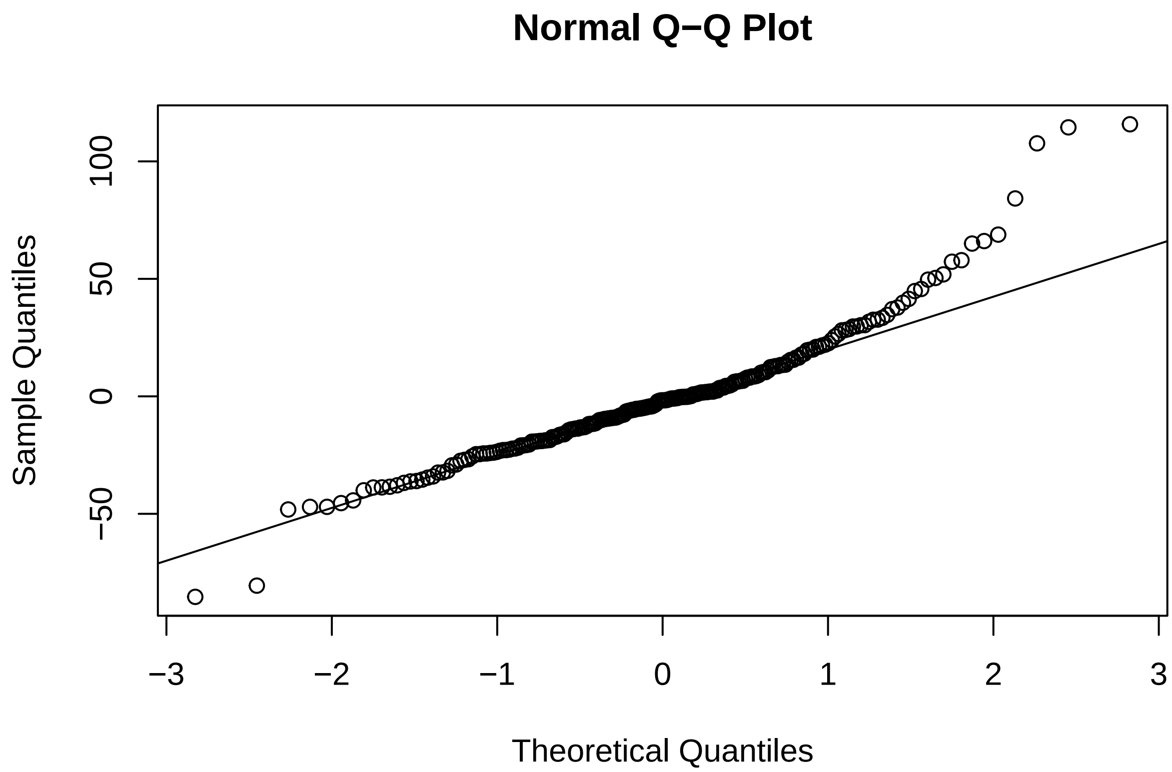
```
MuMIn::r.squaredGLMM(m5.1)
```

```
##            R2m       R2c
## [1,] 0.6899193 0.8175096
```

```
#check assumptions
plot(m5.1)
```



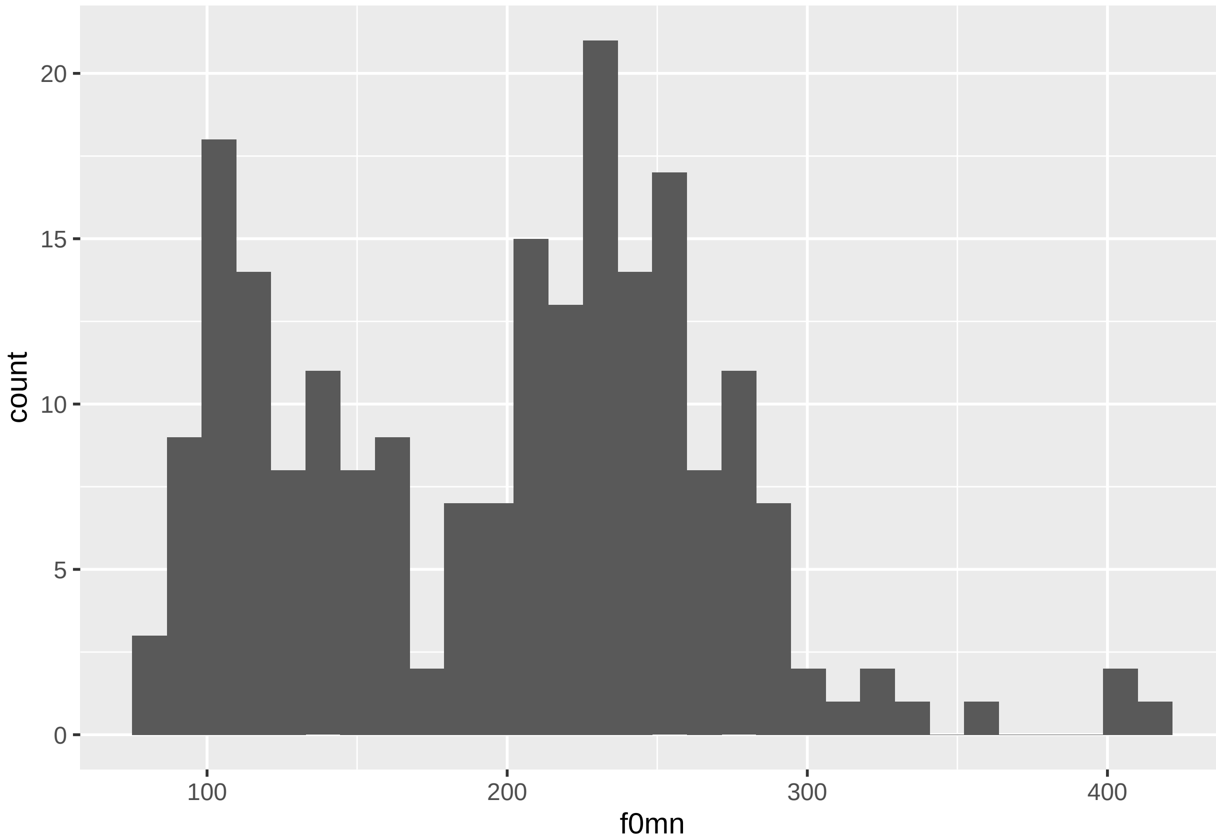

```
qqnorm(resid(m5.1))
qqline(resid(m5.1))
```




```
ggplot(politeness, aes(x= f0mn)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 12 rows containing non-finite values (stat_bin).
```



"We used R (R Core Team, 2019) and lmerTest (Kuznetsova, Brockhoff and Christensen, 2017) to perform a linear mixed effects analysis of the relationship between f0mn, gender and attitude. As random effects, we had intercepts for subjects, and scenario.

Both fixed and random effects accounted for roughly 82% of the variance in the f0mn variable with random effects proportion being 12.7%. Visual inspection shows that both the qqplot and histogram violates the assumption of a mixed effect linear model. The more robust generalized mixed effect model with a link function would be preferred. But as did was not the task such model was not constructed.

f0mn was found to be significantly modulated by gender. $\beta_2 = -115, SE = 12.16, p < 0.05$ Attitude also showed a significant modulating of f0mn $\beta_1 = -14.8, SE = 4, p < 0.05$