practical_exercise_2, Methods 3, 2021, autumn semester

Luke Ring

2021-09-27

Assignment 1: Using mixed effects modelling to model hierarchical data

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`

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

Answers

Exercise 1, part 1 The politeness dataset contains the data obtained from the study of Korean formal and informal speech which investigated the fundamental frequency of male and female participants' speech in a variety of formal and informal scenarios.

The following table describes the variables in the dataset:

Variable	Description
`subject`	participant ID
`gender`	participant's gender
`scenario`	the experimental scenario from 1 to 7 such as "asking a favour"
`attitude`	either 'inf' for informal stimuli or 'pol' for formal stimuli
`total_duration`	duration of participant's response in seconds
`f0mn`	mean fundamental frequency (f0) of the participant's speech
`hiss_count`	number of times the participants made a noisy breath intake

The `gender`, `scenario` and `attitude` variables should be encoded as factors as they are categorical in this dataset and do not have continuous relationship between variable values. In addition, these variables have non-unique values across participants, and are not ordered.

```
# load the data
politeness <- read.csv("politeness.csv")
# Encode attitude and gender as factors
politeness$attitude <- as.factor(politeness$attitude)
politeness$gender <- as.factor(politeness$gender)
# we really want scenario as a factor as well
politeness$scenario <- as.factor(politeness$scenario)
```

```
# Create a subset dataframe for subject F1 only
pf1 <- politeness[politeness$subject == "F1", ]

# make model predicting f0mn by scenario (integer)
m1 <- lm(f0mn ~ as.integer(scenario), data = pf1)
# get model matrix
mm1 <- model.matrix(m1)
# make model predicting f0mn by scenario (factor)
m2 <- lm(f0mn ~ scenario, data = pf1)
# get model matrix
mm2 <- model.matrix(m2)
```

Exercise 1, part 2 Model using scenario as integer

```
summary(m1)
```

```
## 
## Call:
## lm(formula = f0mn ~ as.integer(scenario), data = pf1)
## 
## 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
```

```
mm1
```

```
##    (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
```

Model using scenario as factor

```
summary(m2)
```

```
##
## Call:
## lm(formula = f0mn ~ scenario, data = pf1)
##
## 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
```

```
mm2
```

```
##    (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"
```

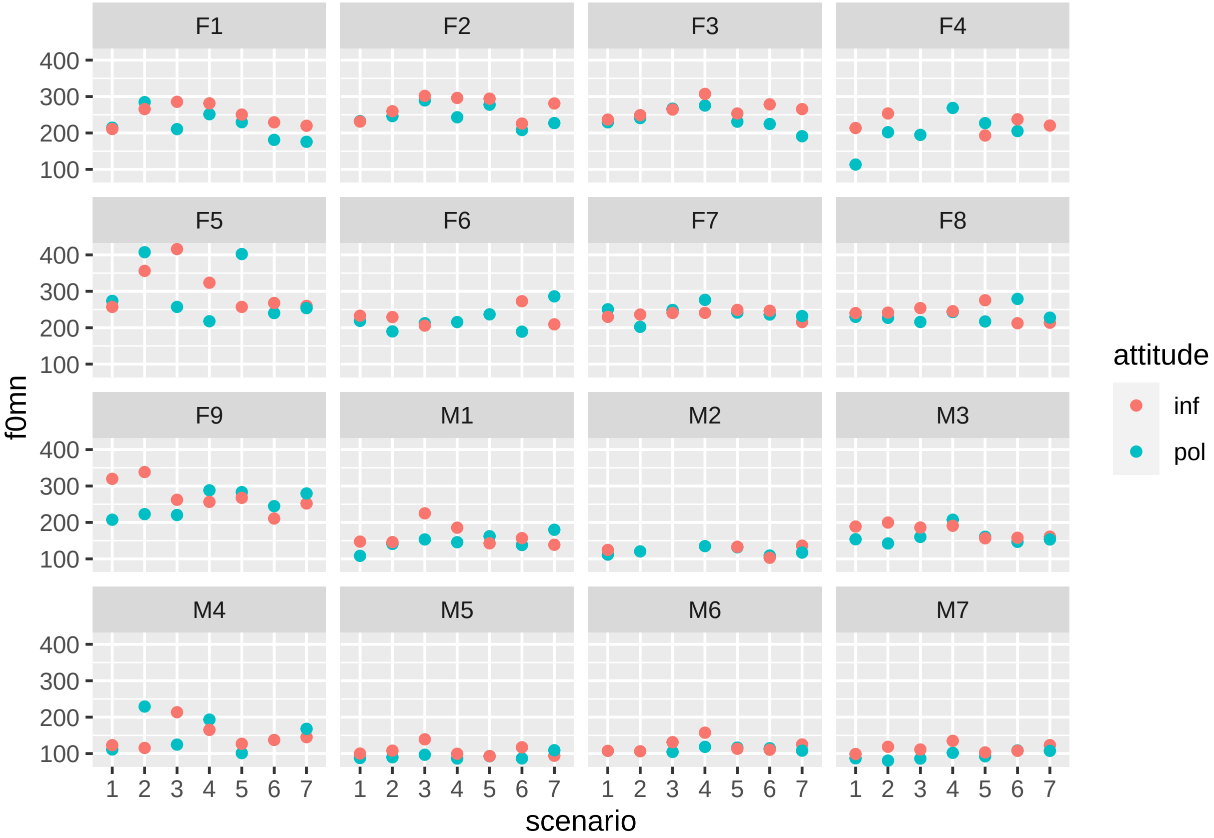
The above output shows the difference in model matrices between scenario encoded as an integer and factor. The integer version treats scenario as a continous variable, whereas the factorised version creates a regression line per scenario.

For this dataset, scenario should be a factor, the scenarios are not a continuous variable and depending on the prescribed scenario, the participants may have a different f0 and this wont consistently increase or decrease across scenarios.

```
# plot participant data by subject and scenario
politeness %>% ggplot(aes(scenario, f0mn, color = attitude)) +
    geom_point() +
    facet_wrap(vars(subject))
```

Exercise 1, part 3

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



We can see that different subjects have different baseline frequencies as well as different within-subject variation between scenarios. Just based on visual examination of the data, it does seem like there is some consistency of the relative f0 changes per scenario across participants.

Exercise 2 - comparison of models

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*
   v. which of the models has the lowest residual standard deviation, also compare the Akaike Information Criterion `AIC`?
   vi. which of the second-level effects explains the most variance?
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_
   ii. build a single-level model that models *f0mn* as dependent on *gender* using this new dataset
   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?)
   iv. Also make a quantile-quantile plot for the residuals of the multilevel model with two intercepts. Does it look alright?
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)

Answers

```
# single level
m3 <- lm(formula = f0mn ~ gender, data = politeness)
# scenario intercept
m4 <- lmer(formula = f0mn ~ gender + (1 | scenario), data = politeness)
# subject intercept
m5 <- lmer(formula = f0mn ~ gender + (1 | subject), data = politeness)
# subject and scenario intercept
m6 <- lmer(formula = f0mn ~ gender +
            (1 | subject) + (1 | scenario), data = politeness)
AIC(m3)
```

Exercise 2, part 1

```
## [1] 2163.971
```

```
deviance(m3)
```

```
## [1] 327033.6
```

```
anova(m4, m5, m6)
```

```
## refitting model(s) with ML (instead of REML)

## Data: politeness
## Models:
## m4: f0mn ~ gender + (1 | scenario)
## m5: f0mn ~ gender + (1 | subject)
## m6: f0mn ~ gender + (1 | subject) + (1 | scenario)
##    npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m4    4 2162.3 2175.7 -1077.1   2154.3
## m5    4 2112.1 2125.5 -1052.0   2104.1 50.2095  0
## m6    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
```

```
piecewiseSEM::rsquared(c(m4, m5, m6))
```

```
##   Response   family     link method  Marginal Conditional
## 1     f0mn gaussian identity   none 0.6779555   0.6967788
## 2     f0mn gaussian identity   none 0.6681651   0.7899229
## 3     f0mn gaussian identity   none 0.6677206   0.8077964
```

```
# Second level variance explained by m4
(piecewiseSEM::rsquared(m4)$Conditional - piecewiseSEM::rsquared(m4)$Marginal)
```

```
## [1] 0.01882337
```

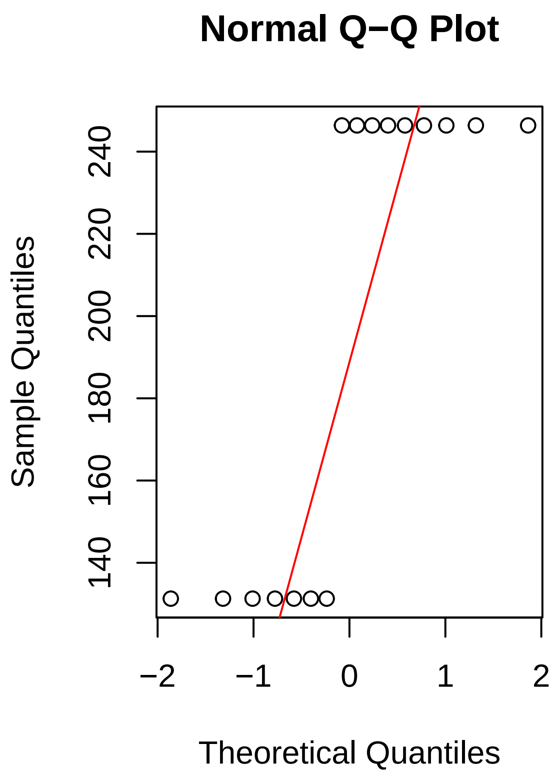
```
# Second level variance explained by m5
(piecewiseSEM::rsquared(m5)$Conditional - piecewiseSEM::rsquared(m5)$Marginal)
```

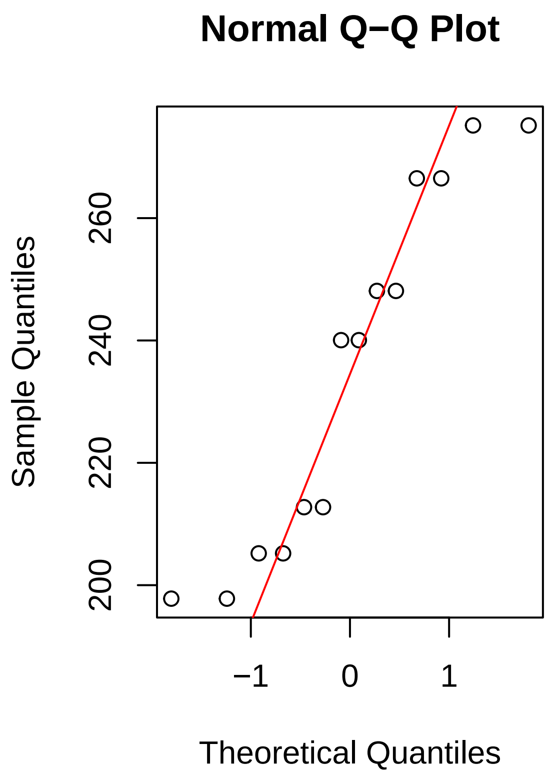
```
## [1] 0.1217578
```

Comparing the above models, we see that the model that has the lowest AIC and deviance is m6, which uses random intercepts for subject and scenario. The single level model performs the worst in both cases, and this makes sense as we do not expect all participants to have the same f0 as their voices have naturally occuring differences (not just based on gender), as well as scenario based differences, neither of which are taken into account from the single level model.

Of the three multi-level models model m6, using random intercepts for subject and scenario has the most explained variance with for the entire model $R^2 \approx 0.81$ or 81%. Model m4 used scenario as the second level, and subtracting the conditional from marginal R^2 gives us 0.012 or 1.2% variance explained by the scenario. For m5, the second level effect used was subject, and using the same method, we get 0.122 or 12% of the variance explained by differences between subjects. As such, subject is the second level effect that explains more variance.

```
# create data frame with aggregated f0mn values
politeness_aggregated <-
    politeness[!is.na(politeness$f0mn), ] %>%
    group_by(subject) %>%
    summarize(subject = subject[1], gender = gender[1], f0mn = mean(f0mn))

# aggregate model
m7 <- lm(f0mn ~ gender, data = politeness_aggregated)

# compare qq plots
par(mfrow = c(1, 2))
qqnorm(fitted.values(m7))
qqline(fitted.values(m7), col = "red")
qqnorm(fitted.values(m2))
qqline(fitted.values(m2), col = "red")
```





Exercise 2, part 2

```
par(mfrow = c(1, 1))
```

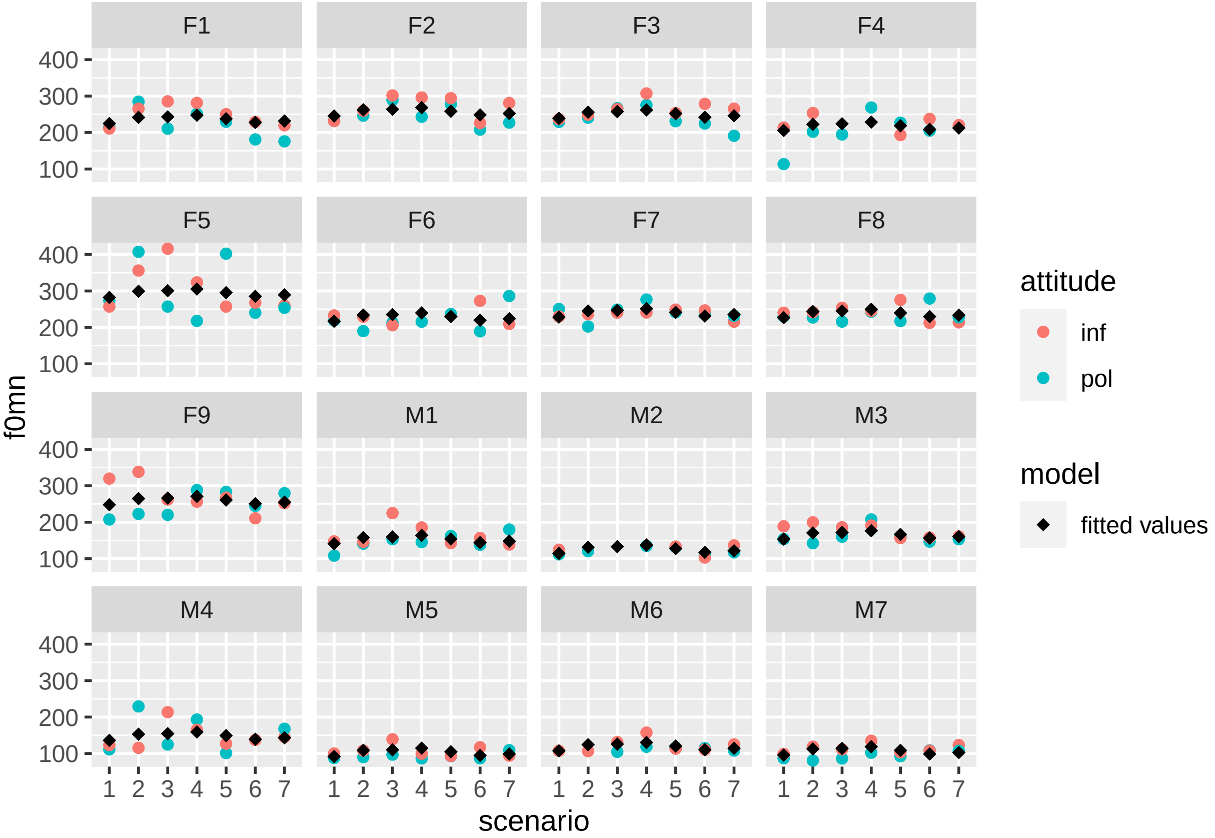
Assessing the QQ-plots of the single-level models it seems that the aggregated model m7's residuals are worse off than those of model m2. The residuals of model m2 are better dispersed along the line - however it still doesn't look fantastic. The QQ-plot of the multilevel model m6 looks better than any of the single level ones, with data points closer to the line and more evenly dispersed on both sides of the line.

```
ff <- fixef(m6)
rf <- ranef(m6)
rf <- as.data.frame(rf)

# yes this is a janky way to do it
# but we manually calculate predictions
# using the fixed and random effects
# from the model
politeness$effect_gender <- 0.0
politeness[politeness$gender == "F", ]$effect_gender <- ff[1]
politeness[politeness$gender == "M", ]$effect_gender <- ff[1] + ff[2]

# join subject random effects
politeness$intercept_subject <-
    left_join(politeness, rf,
    by = c("subject" = "grp"), copy = TRUE, keep = FALSE)$condval
# join scenario random effects
politeness$intercept_scenario <-
    left_join(politeness, rf,
    by = c("scenario" = "grp"), copy = TRUE, keep = FALSE)$condval
```

```
# calculate model fitted values per data point
politeness$predicted <-
    politeness$effect_gender +
    politeness$intercept_subject +
    politeness$intercept_scenario

# plot subject data and fitted values
politeness %>% ggplot(aes(scenario, f0mn, color = attitude)) +
    geom_point() +
    geom_point(aes(y = predicted, shape = "fitted values"),
        color = "black", size = 2) +
    scale_shape_manual(name = "model", values = c(18)) +
    facet_wrap(vars(subject))
```

Exercise 2, part 3

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



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

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.

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

Answers

```
# model attitude as additional fixed effect
m8 <- lmer(formula = f0mn ~ gender + attitude +
    (1 | subject) + (1 | scenario), data = politeness)

# now add interaction effect between gender and attitude
m9 <- lmer(formula = f0mn ~ gender + attitude + gender:attitude +
    (1 | subject) + (1 | scenario), data = politeness)

summary(m9)
```

Exercise 3, part 1

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f0mn ~ gender + attitude + gender:attitude + (1 | subject) +
##     (1 | scenario)
##    Data: politeness
##
## REML criterion at convergence: 2058.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8120 -0.5884 -0.0645  0.4014  3.9100
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 584.4    24.17
##  scenario (Intercept) 106.4    10.32
##  Residual             885.5    29.76
## Number of obs: 212, groups:  subject, 16; scenario, 7
##
## Fixed effects:
##                       Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)            255.618      9.761   20.909  26.186  < 2e-16 ***
## genderM               -118.232     13.531   17.158  -8.738    1e-07 ***
## attitudepol          -17.192      5.423  188.445  -3.170  0.00178 **
## genderM:attitudepol    5.544      8.284  188.491   0.669  0.50412
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM atttdp
## genderM     -0.606
## attitudepol -0.286  0.206
## gndrM:tttdp  0.187 -0.309 -0.654
```

The model m9 can be read as following: The intercept for women/inf are $\approx 256Hz$, when we look at men with the same attitude their pitch drops by $\approx 118Hz$. Overall a polite attitude will result in a drop in pitch by $\approx 17Hz$, however for men it will only be $-17.2 + 5.5 \approx 11.6Hz$. Korean women's relative drop in pitch is therefore larger than male's in a polite situation according to this sample.

```r
# model comparison showing AIC and variance
anova(m6, m8, m9)
```

Exercise 3, part 2

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: politeness
## Models:
## m6: f0mn ~ gender + (1 | subject) + (1 | scenario)
## m8: f0mn ~ gender + attitude + (1 | subject) + (1 | scenario)
## m9: f0mn ~ gender + attitude + gender:attitude + (1 | subject) + (1 | scenario)
##    npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## m6    5 2105.2 2122.0 -1047.6   2095.2
## m8    6 2094.5 2114.6 -1041.2   2082.5 12.6868  1  0.0003683 ***
## m9    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
```

```r
# m6 rsd
sigma(m6)
```

```
## [1] 30.65803
```

```r
# m7 rsd
sigma(m8)
```

```
## [1] 29.71087
```

```r
# m8 rsd
sigma(m9)
```

```
## [1] 29.75684
```

Based on the above comparisons, we can see that m8 has the lowest AIC score, m9 has very slightly less (0.5) residual variance, although overall it appears m8 is a decent model that probably isn't overfitted. m8 also has slightly less residual standard deviation than the other two models.

Exercise 3, part 3 This dataset consists of the basic demographic information of 16 Korean participants, and their observed pitch in different situations with either an informal or polite attitude.

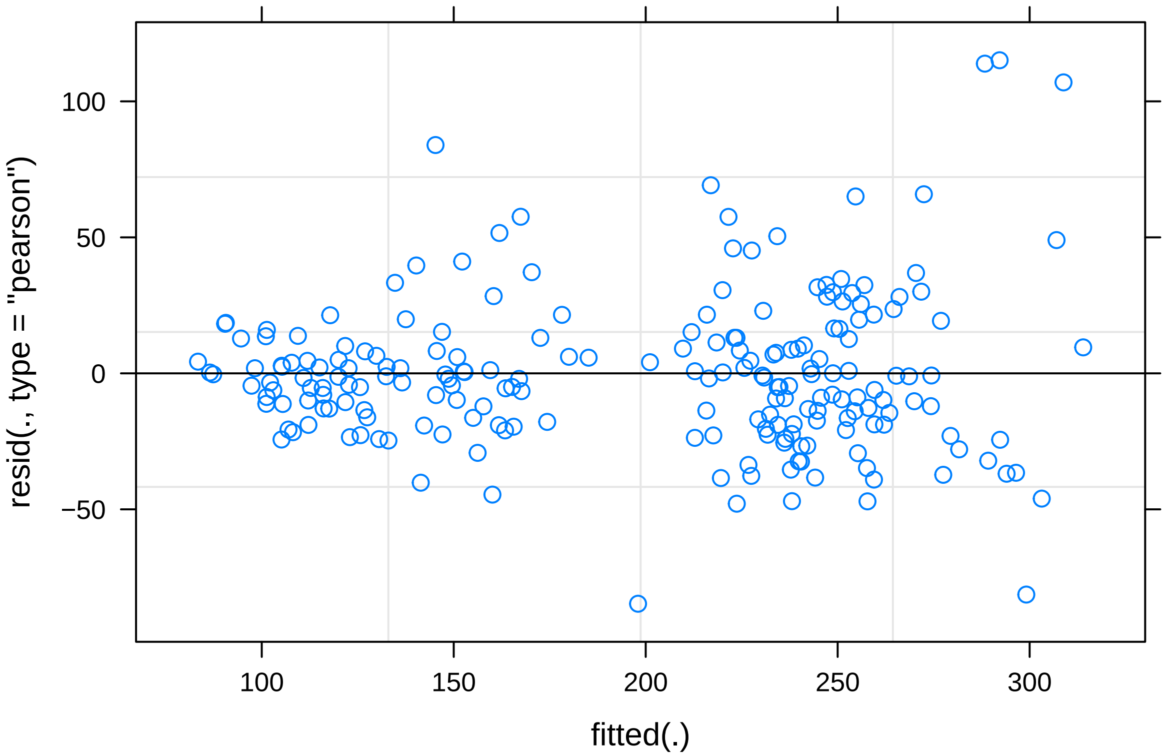
```
summary(m8)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: f0mn ~ gender + attitude + (1 | subject) + (1 | scenario)
##    Data: politeness
##
## REML criterion at convergence: 2065.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8511 -0.6081 -0.0602  0.4329  3.8745
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 585.6    24.20
##  scenario (Intercept) 106.7    10.33
##  Residual             882.7    29.71
## Number of obs: 212, groups:  subject, 16; scenario, 7
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  254.398      9.597   19.480  26.507  < 2e-16 ***
## genderM     -115.437     12.881   14.066  -8.962 3.44e-07 ***
## attitudepol  -14.819      4.096  189.652  -3.618 0.000381 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) gendrM
## genderM     -0.587
## attitudepol -0.220  0.006
```

```
anova(m8)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##          Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
## gender    70896   70896     1  14.066  80.314 3.439e-07 ***
## attitude  11554   11554     1 189.652  13.089 0.0003808 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(m8)
```



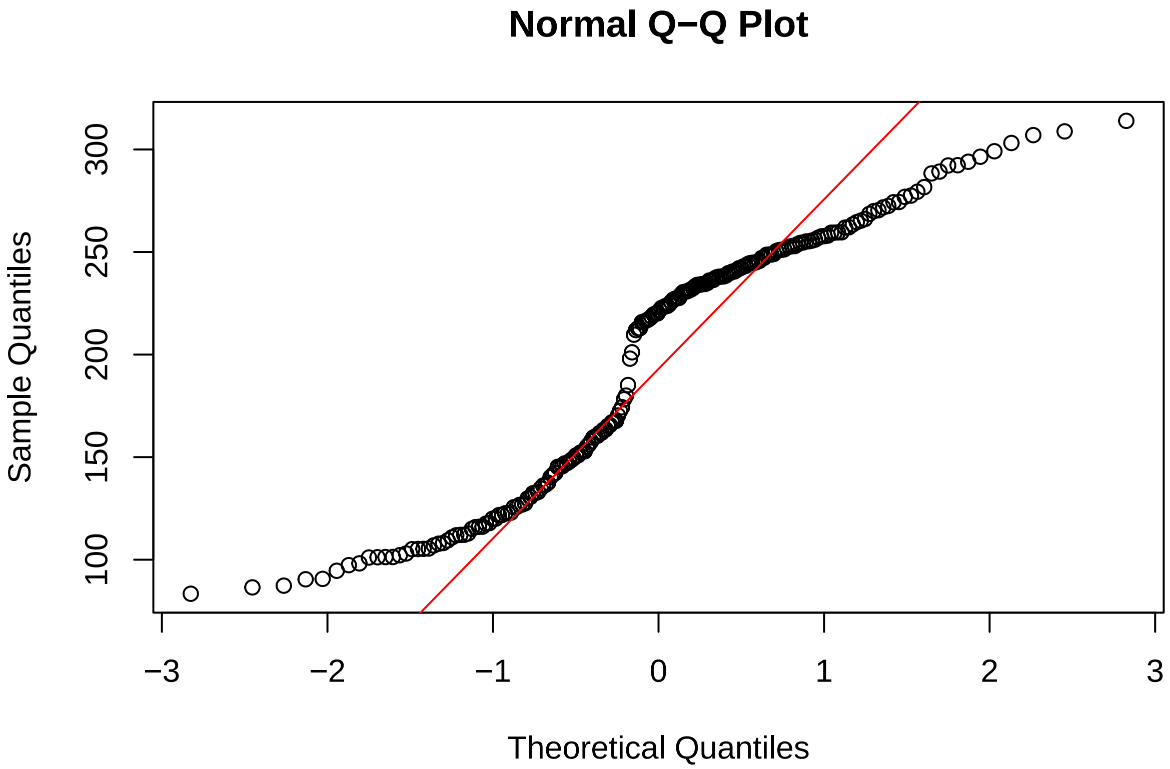
Non-surprisingly our model showed that women on average have a higher pitch than men BUT it also suggested a negative relationship between *attitude* and pitch with p-values<0.001. It would seem that both Korean men and women frequency of voice drops when having a polite attitude.

Subjects and scenarios should have different intercepts because it would be assumed they would all have different baselines, (and therefore need different intercepts to account for this). Different subjects will naturally already speak at a different pitch level, so separate intercepts allows us to account for these differences. The different scenarios may also need separate intercepts as certain scenarios may result in participants lowering or raising their pitch to meet the appropriate ambiance of the scenario. Once again, by including separate intercepts for scenarios then we should be accounting for these differences in our models.

Furthermore the output of the summary function shows that more variance is explained by the random effect of the subject than that of the scenario, further strengthening the choice of multilevel modelling.

And here's a QQ-plot of our chosen model

```
qqnorm(fitted.values(m8))
qqline(fitted.values(m8), col = "red")
```


Normal Q−Q Plot
Sample Quantiles
100
150
200
250
300
−3
−2
−1
0
1
2
3
Theoretical Quantiles