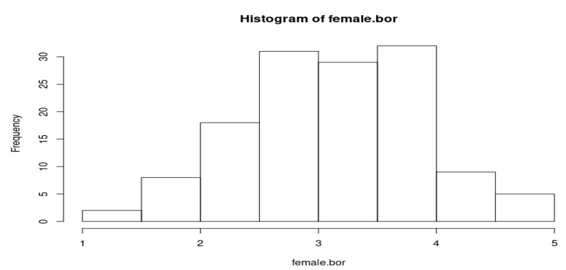
Lab assignment 3: generalized linear mixed-effects models and selfies

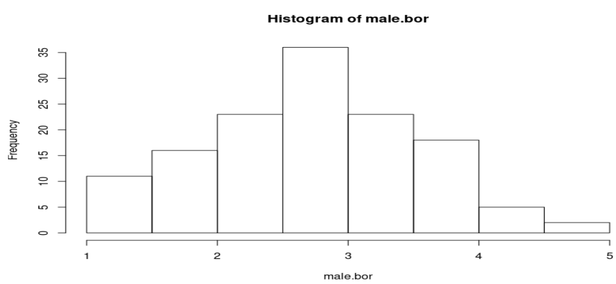
Hunter Sterk, Dirk Keller, Markos Polos

30-03-2021

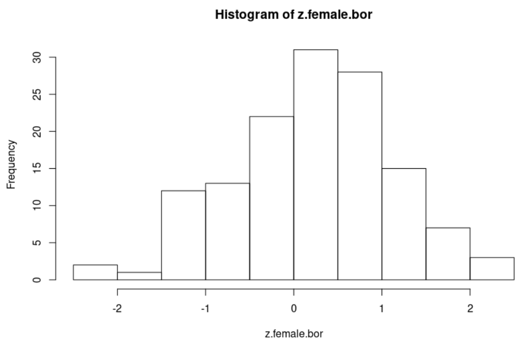
Q1: Does the gender of the selfie-taker predict boringness responses?

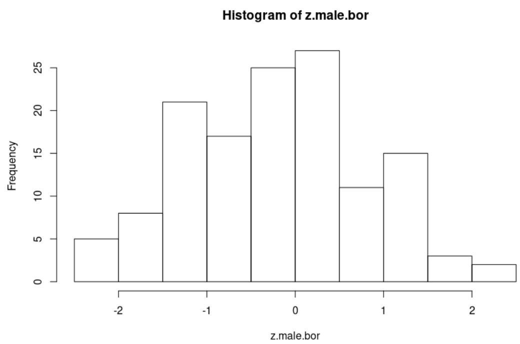
```
library(dplyr)
library(ggplot2)
library(lme4)

selfies <- read.csv("C:/Users/DK/Desktop/Experimentation in Psychology
and Linguistics/Lab 3 Mixed design/selfies.csv")
str(selfies)

# eliminate trials with missing values
c.selfies <- na.omit(selfies)
nrow(selfies)-nrow(c.selfies)

# prepare data set for female/male condition of boring measures
self.gen.bor <- c.selfies %>% group_by(ResponseId, StimGender) %>%
summarise(Boring = mean(Boring))
female.bor <- self.gen.bor$Boring[self.gen.bor$StimGender == 'Female']
male.bor <-  self.gen.bor$Boring[self.gen.bor$StimGender == 'Male']

# check assumptions
shapiro.test(female.bor)
shapiro.test(male.bor)
hist(female.bor)
hist(male.bor)
```





```
# transform data into zscores
Zscore = (self.gen.bor$Boring -
mean(self.gen.bor$Boring))/sd(self.gen.bor$Boring)
z.self.gen.bor <- cbind(data.frame(self.gen.bor), data.frame(Zscore))

z.female.bor <- z.self.gen.bor$Zscore[z.self.gen.bor$StimGender ==
'Female']
z.male.bor <-  z.self.gen.bor$Zscore[z.self.gen.bor$StimGender ==
'Male']

hist(z.female.bor)
hist(z.male.bor)

# significance test (parametric)
t.test(z.female.bor, z.male.bor, paired = TRUE, var.equal = TRUE)
```






```
                                                                       #
Boring categorization,
BoringYesNo <- selfies$Boring
selfies <- cbind(selfies, data.frame(BoringYesNo))
selfies$BoringYesNo[selfies$Boring < 3] <- 0
selfies$BoringYesNo[selfies$Boring > 3] <- 1
selfies$BoringYesNo[selfies$Boring == 3] <- NaN

c.selfies <- na.omit(selfies)

# sum contrast coding (different values)
#conStimGender <- c.selfies$StimGender
#c.selfies <- cbind(c.selfies, data.frame(conStimGender))
#c.selfies$conStimGender[c.selfies$StimGender == 'Female'] <- 1
#c.selfies$conStimGender[c.selfies$StimGender == 'Male'] <- -1

m1 <- glmer(BoringYesNo ~ 1 + (1 | ResponseId), c.selfies, family =
binomial(link = "logit"))
```

```
summary(m1)
m2 <- glmer(BoringYesNo ~ 1 + StimGender + (1 |ResponseId), c.selfies,
family = binomial(link = "logit"))
summary(m2)
m3 <- glmer(BoringYesNo ~ 1 + StimGender + (1 + StimGender |ResponseId),
c.selfies, family = binomial(link = "logit"))
summary(m3)

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

According to the significant amount of variance explained, model 2 was chosen. Moreover, the results are consistent with the paired t-test, both yield a significant effect of stimulus gender.

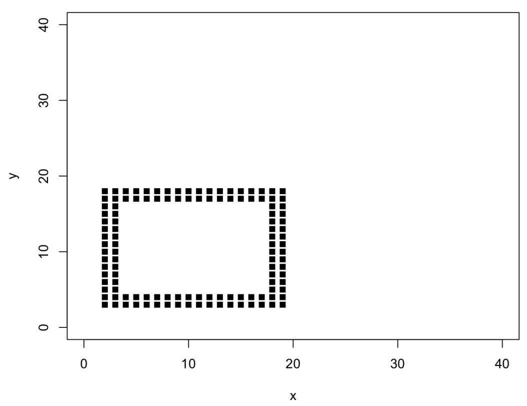
Q2: Reasoning about the model

T-test do not account or correct for dependencies of subjects in the data, or other confounding variables that affect the measures in a dependent manner. In addition, to perform t-tests data is aggregated hence much information is lost and relevant trends might be hidden in the aggregated data (e.g. Simon paradoxon). Similar to the t-test, in the linear model, other variables might be theoretical relevant (e.g. interaction effects, additional factors co-influence the measurement) or might be unaccounted observed patterns of 'random variables' that affect comparison hence the validity of interpretation. Examples in the present data set are participant gender age or type of social media platform (e.g. dating website vs. twitter). An additional problem lies in the binary transformation of the measurement that gives a lot of information away (granularity, orderedness etc.). In summary, the generalized linear model improves on the t-test with respect to confound controls, takes dependencies into account and does not rely on too coarse aggregated data. Alltogether might affect the significance value and the validity of the interpretation. However, both models may be still a too strong oversimplification, that does not address other relevant factors (contributing to the construct= and possible confounding variables .

Q3: Predictions of the model

```
novel_selfies <- read.csv("C:/Users/DK/Desktop/Experimentation in
```

```
Psychology and Linguistics/Lab 3 Mixed design/novel_data.csv")
c.novel_selfies <- na.omit(novel_selfies)

m4 <- glmer(BoringYesNo ~ 1 + StimGender * Gender + (1 + StimGender |
ResponseId), c.selfies, family = binomial(link = "logit"))
print(summary(m4))

ResponseId <- c(1:400)
novel_selfies <- cbind(novel_selfies, data.frame(ResponseId))
prob.GeStim <- predict(m4, newdata = novel_selfies, allow.new.levels =
TRUE)

drawprobabilities <- function(probs) {
  if (length(probs) != 400) {
      print("Wrong length of the vector of calculated probabilities.
Should be 400 data points.")
  } else {
      matrixprobs <- matrix(ifelse(probs > 0.5, "X", ""), nrow = 20)
      x <- rep(NA, 400)
      y <- rep(NA, 400)
      k <- 1
      for (i in 1:20) {
         for (j in 1:20) {
             if (matrixprobs[i, j] == "X") {
                 y[k] <- i
                 x[k] <- j
                 k <- k + 1

             }
         }
      }
      plot(x, y, xlim = c(0, 40), ylim = c(0, 40), pch = 15)
  }
}
drawprobabilities(prob.GeStim)

rm(m4, prob.GeStim, ResponseId,drawprobabilities)
```



Q4: Ordinal model

```
library(ordinal) #install.packages('ordinal')
selfies <- read.csv("C:/Users/DK/Desktop/Experimentation in Psychology
and Linguistics/Lab 3 Mixed design/selfies.csv")
c.selfies <- na.omit(selfies)

# model comparison# model comparison
m1 <- clmm(factor(Boring) ~ 1 + StimGender + (1 | ResponseId), link =
"probit", c.selfies)
summary(m1)
m2 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId), link = "probit", c.selfies)
summary(m2)

m3 <- clmm(factor(Boring) ~ 1 + StimGender * Gender + (1 + StimGender |
ResponseId), link = "probit", c.selfies)
summary(m3)
m4 <- clmm(factor(Boring) ~ 1 + StimGender * Gender + (1 + StimGender *
Gender | ResponseId), link = "probit", c.selfies)
summary(m4)

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

Model 2 is the best model here, interaction between StimGender and Gender does not explain more variance.

```
# Additional confounder with more than 3 levels
m0.1 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId) + (1 + StimGender | Dur), link = "probit", c.selfies) #*
m0.2 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId) + (1 + StimGender | Age), link = "probit", c.selfies)
m0.3 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId) + (1 + StimGender | Country), link = "probit", c.selfies)
m0.4 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId) + (1 + StimGender | Socialmedia), link = "probit",
c.selfies)
m0.5 <- clmm(factor(Boring) ~ 1 + StimGender + (1 + StimGender |
ResponseId) + (1 + StimGender | Selfietaking), link = "probit",
c.selfies)
summary(m0.1)
summary(m0.2)
```

```
summary(m0.3)
summary(m0.4)
summary(m0.5)
anova(m2, m0.1)
anova(m2, m0.2)
anova(m2, m0.3)
anova(m2, m0.4) # *for fixed effects
anova(m2, m0.5)

# Including them as fixed effects
m0.1 <- clmm(factor(Boring) ~ 1 + StimGender * Dur + (1 + StimGender |
ResponseId), link = "probit", c.selfies)
m0.2 <- clmm(factor(Boring) ~ 1 + StimGender * Age + (1 + StimGender |
ResponseId), link = "probit", c.selfies) # interaction*
m0.3 <- clmm(factor(Boring) ~ 1 + StimGender * Country + (1 + StimGender
| ResponseId), link = "probit", c.selfies)
m0.4 <- clmm(factor(Boring) ~ 1 + StimGender * Socialmedia + (1 +
StimGender | ResponseId), link = "probit", c.selfies)
m0.5 <- clmm(factor(Boring) ~ 1 + StimGender * Selfietaking + (1 +
StimGender | ResponseId), link = "probit", c.selfies)

# Additional confounder with less than 3 levels
m0.6 <- clmm(factor(Boring) ~ 1 + StimGender * Gender + (1 + StimGender
| ResponseId), link = "probit", c.selfies)
m0.7 <- clmm(factor(Boring) ~ 1 + StimGender * Tilt + (1 + StimGender |
ResponseId), link = "probit", c.selfies) # **StimGender, *interaction
effect
m0.8 <- clmm(factor(Boring) ~ 1 + StimGender * Distance + (1 +
StimGender | ResponseId), link = "probit", c.selfies) # ** Distance,
***interaction effect
m0.9 <- clmm(factor(Boring) ~ 1 + StimGender * Eyes + (1 + StimGender |
ResponseId), link = "probit", c.selfies) # ***StimGender

summary(m0.5)
summary(m0.6)
summary(m0.7)
summary(m0.8)
summary(m0.9)
anova(m2, m0.5, m0.6, m0.7, m0.8, m0.9)
```

When accounting for possible confounding variables with more than 3 levels, StimGender is still significant in all 4 models and the models with a more complex random structure are not able to significantly account for more variance than the baseline model (m2)

When including possible confounds as random effects, the significance of StimGender is not affected. However, when including the less than 3 level variables as fixed effects StimGender is often significant. Only for Distance (and Age, Selfietaking, Country) the effect of StimGender was not significant anymore. Importantly the interaction effect of Tilt and Distance was significant. Importantly the model including tilt as fixed effect explains more variance than the baseline model and has an significant interaction effect. Although, generally speaking many possible confounds had no effect on the difference of male selfies from female ones with respect to boringness and many? However, Distance and especially Tilt might be contribute to the difference hence the difference might be restricted and driven by these specific factors.