Data Analyses

Kaleem Ullah

2024-04-22

Exploratory Data Analyses

```r
df <- read.csv("new_csv/preprocessed_data.csv")
length(unique(df$ppn))
```

```
## [1] 1405
```

Preliminary Analyses

```r
df$culture <- as.factor(df$culture)
# df$ppn <- as.factor(df$ppn)

# Hypothesis 1
accuracy_fit_1 <- glmer(accuracy ~ culture + (1 | ppn2),
                        data = df,
                        family = binomial,
                        control = glmerControl(optimizer = "bobyqa"),
                        nAGQ = 10)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
accuracy_fit_2 <- glmer(accuracy ~ culture + (1 | ppn2/emotion),
                        data = df,
                        family = binomial,
                        control = glmerControl(optimizer = "bobyqa"))
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(accuracy_fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
##  Family: binomial  ( logit )
## Formula: accuracy ~ culture + (1 | ppn2)
##    Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  31235.3  31269.0 -15613.7  31227.3    33722
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.4857 -0.4857 -0.4488 -0.4459  2.2429
```

```
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ppn2   (Intercept) 0        0
## Number of obs: 33726, groups:  ppn2, 1405
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.61554    0.02536 -63.705  < 2e-16 ***
## cultureCN    0.17131    0.03491   4.906 9.27e-07 ***
## cultureNL    0.01323    0.03578   0.370    0.712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN
## cultureCN -0.726
## cultureNL -0.709  0.515
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```r
summary(accuracy_fit_2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: accuracy ~ culture + (1 | ppn2/emotion)
##    Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  29033.1  29075.2 -14511.6  29023.1    33721
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9768 -0.2436 -0.2177 -0.2157  2.2118
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  emotion:ppn2 (Intercept) 3.486    1.867
##  ppn2         (Intercept) 0.000    0.000
## Number of obs: 33726, groups:  emotion:ppn2, 11240; ppn2, 1405
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.55955    0.04894 -52.295  < 2e-16 ***
## cultureCN    0.24295    0.04162   5.838 5.28e-09 ***
## cultureNL    0.01861    0.04246   0.438    0.661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN
## cultureCN -0.482
## cultureNL -0.438  0.512
```

```
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```r
anova(accuracy_fit_1, accuracy_fit_2)
```

```
## Data: df
## Models:
## accuracy_fit_1: accuracy ~ culture + (1 | ppn2)
## accuracy_fit_2: accuracy ~ culture + (1 | ppn2/emotion)
##                npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## accuracy_fit_1    4 31235 31269 -15614    31227
## accuracy_fit_2    5 29033 29075 -14512    29023 2204.2  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Likert 1

```r
# Hypothesis 2
iden_fit <- lmer(likert1 ~ culture + (1 | ppn),
                 data=df)
summary(iden_fit)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: likert1 ~ culture + (1 | ppn)
##    Data: df
##
## REML criterion at convergence: 145216.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2314 -0.6130 -0.0154  0.5916  4.5092
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ppn      (Intercept) 3.810    1.952
##  Residual             4.261    2.064
## Number of obs: 32860, groups:  ppn, 1401
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  4.78238    0.05577  85.753
## cultureCN    0.13068    0.02791   4.682
## cultureNL    0.18045    0.02790   6.467
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN
## cultureCN -0.250
## cultureNL -0.250  0.500
```

Likert 2

```r
iden_fit <- lmer(likert2 ~ culture + (1 | ppn),
                 data=df)
```

```r
summary(iden_fit)$coeff
```

```
##             Estimate Std. Error   t value
## (Intercept) 5.0384029 0.05583624 90.235346
## cultureCN   0.1139401 0.02933497  3.884106
## cultureNL   0.3045634 0.02931788 10.388315
```

```r
# anova(iden)
```

```r
# Hypothesis 3
glm_fit <- glmer(accuracy ~ culture + likert1 + (1 | ppn),
                 data = df,
                 family = binomial,
                 control = glmerControl(optimizer = "bobyqa"),
                 nAGQ = 10)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
glm_fit_2 <- glmer(accuracy ~ culture + likert1 + (1 | ppn),
                 data = df,
                 family = binomial)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(glm_fit)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
##  Family: binomial  ( logit )
## Formula: accuracy ~ culture + likert1 + (1 | ppn)
##    Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  30503.9  30545.9 -15247.0  30493.9    32855
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -0.5259 -0.4736 -0.4528 -0.4243  2.4060
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ppn    (Intercept) 0        0
## Number of obs: 32860, groups:  ppn, 1401
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.755947   0.036076 -48.674  < 2e-16 ***
## cultureCN    0.171314   0.035323   4.850 1.23e-06 ***
## cultureNL    0.011224   0.036196   0.310    0.757
## likert1      0.029923   0.005147   5.814 6.10e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN cltrNL
```

```
## cultureCN -0.504
## cultureNL -0.486  0.515
## likert1   -0.703 -0.017 -0.025
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

summary(glm_fit_2)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
## Family: binomial  ( logit )
## Formula: accuracy ~ culture + likert1 + (1 | ppn)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##  30503.9  30545.9 -15247.0  30493.9    32855
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -0.5259 -0.4736 -0.4528 -0.4243  2.4060
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ppn    (Intercept) 4e-14    2e-07
## Number of obs: 32860, groups:  ppn, 1401
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.755947   0.036066 -48.687  < 2e-16 ***
## cultureCN    0.171314   0.035318   4.851 1.23e-06 ***
## cultureNL    0.011224   0.036191   0.310    0.756
## likert1      0.029923   0.005146   5.815 6.08e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN cltrNL
## cultureCN -0.504
## cultureNL -0.486  0.515
## likert1   -0.703 -0.017 -0.025
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Demographics

```
demo_data <- read.csv("new_csv/prelim_demo.csv", header=T)
```

```
ratio <- sum(demo_data$gender == ' ') / sum(demo_data$gender == ' ')
```

Mean Age: NA Ratio of Male/Female: 1.1007519

Preliminary Analyses

```r
# This R file contains code to read and analyze preprocessed data from Gorilla
df$culture <- as.factor(df$culture)
df$ppn <- as.factor(df$ppn)

# Hypothesis 1
accuracy_fit_1 <- glmer(accuracy ~ culture + (1 | ppn2),
                        data = df,
                        family = binomial,
                        control = glmerControl(optimizer = "bobyqa"),
                        nAGQ = 10)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
accuracy_fit_2 <- glmer(accuracy ~ culture + (1 | ppn2/emotion),
                        data = df,
                        family = binomial,
                        control = glmerControl(optimizer = "bobyqa"))
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(accuracy_fit_1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 10) [glmerMod]
##  Family: binomial  ( logit )
## Formula: accuracy ~ culture + (1 | ppn2)
##    Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  31235.3  31269.0 -15613.7  31227.3    33722
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.4857 -0.4857 -0.4488 -0.4459  2.2429
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ppn2   (Intercept) 0        0
## Number of obs: 33726, groups:  ppn2, 1405
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.61554    0.02536 -63.705  < 2e-16 ***
## cultureCN    0.17131    0.03491   4.906 9.27e-07 ***
## cultureNL    0.01323    0.03578   0.370    0.712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN
## cultureCN -0.726
## cultureNL -0.709  0.515
## optimizer (bobyqa) convergence code: 0 (OK)
```

```
## boundary (singular) fit: see help('isSingular')
```

`summary(accuracy_fit_2)`

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: accuracy ~ culture + (1 | ppn2/emotion)
##    Data: df
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##  29033.1  29075.2 -14511.6  29023.1    33721
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9768 -0.2436 -0.2177 -0.2157  2.2118
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  emotion:ppn2 (Intercept) 3.486    1.867
##  ppn2         (Intercept) 0.000    0.000
## Number of obs: 33726, groups:  emotion:ppn2, 11240; ppn2, 1405
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.55955    0.04894 -52.295  < 2e-16 ***
## cultureCN    0.24295    0.04162   5.838 5.28e-09 ***
## cultureNL    0.01861    0.04246   0.438    0.661
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) cltrCN
## cultureCN -0.482
## cultureNL -0.438  0.512
## optimizer (bobyqa) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

`anova(accuracy_fit_1, accuracy_fit_2)`

```
## Data: df
## Models:
## accuracy_fit_1: accuracy ~ culture + (1 | ppn2)
## accuracy_fit_2: accuracy ~ culture + (1 | ppn2/emotion)
##                npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## accuracy_fit_1    4 31235 31269 -15614    31227
## accuracy_fit_2    5 29033 29075 -14512    29023 2204.2  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```