Numbers-Experment-MML

Behavioural analysis

Swarnendu Moitra

Table of contents

Analysis based on Persons experiment: Middleton and Moitra 2023

```
library("tidyverse")
library("plotrix")
library("lme4")
library("lmerTest")
library("report")
library("Hmisc")
library("multcomp")
library("emmeans")
source('ggplot_theme_Publication-2.R')
```

```
# Load data
df_slet_1 <- read_csv("data/N72/slet.csv")
df_98zo_1 <- read_csv("data/N72/98zo.csv")
df_goty_1 <- read_csv("data/N72/goty.csv")
df_hebw_1 <- read_csv("data/N72/hebw.csv")
df_ajpl_1 <- read_csv("data/N72/ajpl.csv")
df_ld3j_1 <- read_csv("data/N72/ld3j.csv")

df_slet_2 <- read_csv("data/Rest/slet.csv")
df_98zo_2 <- read_csv("data/Rest/98zo.csv")
df_goty_2 <- read_csv("data/Rest/goty.csv")
df_hebw_2 <- read_csv("data/Rest/hebw.csv")
df_ajpl_2 <- read_csv("data/Rest/ajpl.csv")
df_ld3j_2 <- read_csv("data/Rest/ld3j.csv")
```

```
df_slet_3 <- read_csv("data/Rest_new/slet.csv")
df_98zo_3 <- read_csv("data/Rest_new/98zo.csv")
df_goty_3 <- read_csv("data/Rest_new/goty.csv")
df_hebw_3 <- read_csv("data/Rest_new/hebw.csv")
df_ajpl_3 <- read_csv("data/Rest_new/ajpl.csv")
df_ld3j_3 <- read_csv("data/Rest_new/ld3j.csv")

# Custom Function
data_cleaning <- function(arg1){
arg1 <- arg1
arg1 <- arg1 %>%
        dplyr::select(`Participant Private ID`,`Trial Number`,'Tree Node Key', 'Reaction Tim
        rename(Subject = `Participant Private ID`,
                      Item = `Trial Number`,
                      Condition = 'Task Name',
                      RT = 'Reaction Time',
                      Accuracy = Correct) %>%
        filter(Screen == "Testing",
               `Response Type` == "response") %>%
        mutate(LogRT = log(RT),
        Condition = gsub(" B Testing", "", Condition))
return(arg1)
}

## Use the custom function to clean the data

df_slet_1 <- data_cleaning(df_slet_1)
df_98zo_1 <- data_cleaning(df_98zo_1)
df_goty_1 <- data_cleaning(df_goty_1)
df_hebw_1 <- data_cleaning(df_hebw_1)
df_ajpl_1 <- data_cleaning(df_ajpl_1)
df_ld3j_1 <- data_cleaning(df_ld3j_1)

df_slet_2 <- data_cleaning(df_slet_2)
df_98zo_2 <- data_cleaning(df_98zo_2)
df_goty_2 <- data_cleaning(df_goty_2)
df_hebw_2 <- data_cleaning(df_hebw_2)
df_ajpl_2 <- data_cleaning(df_ajpl_2)
df_ld3j_2 <- data_cleaning(df_ld3j_2)

df_slet_3 <- data_cleaning(df_slet_3)
df_98zo_3 <- data_cleaning(df_98zo_3)
```

```
df_goty_3 <- data_cleaning(df_goty_3)
df_hebw_3 <- data_cleaning(df_hebw_3)
df_ajpl_3 <- data_cleaning(df_ajpl_3)
df_ld3j_3 <- data_cleaning(df_ld3j_3)

data <- rbind(df_slet_1,df_98zo_1,df_goty_1,df_hebw_1,df_ajpl_1,df_ld3j_1,
              df_slet_2,df_98zo_2,df_goty_2,df_hebw_2,df_ajpl_2,df_ld3j_2,
              df_slet_3,df_98zo_3,df_goty_3,df_hebw_3,df_ajpl_3,df_ld3j_3
              )

## Create the summary
data_group <- data %>% group_by(Condition) %>% summarise(RT_mean=mean(RT),RT_SE=std.error(RT)

Data <- data
summary(Data)
```

```
    Subject              Item        Tree Node Key             RT          
 Min.   : 9898448   Min.   : 1.00   Length:19392       Min.   :    16.63  
 1st Qu.: 9992482   1st Qu.:12.75   Class :character   1st Qu.:  1793.88  
 Median :10004852   Median :24.50   Mode  :character   Median :  3318.40  
 Mean   :10383072   Mean   :24.50                      Mean   :  5322.75  
 3rd Qu.:10847984   3rd Qu.:36.25                      3rd Qu.:  6334.27  
 Max.   :10990139   Max.   :48.00                      Max.   :205295.10  
    Accuracy       Spreadsheet: display    Screen            Condition        
 Min.   :0.0000   Length:19392          Length:19392       Length:19392      
 1st Qu.:0.0000   Class :character      Class :character   Class :character  
 Median :1.0000   Mode  :character      Mode  :character   Mode  :character  
 Mean   :0.6358                                                               
 3rd Qu.:1.0000                                                               
 Max.   :1.0000                                                               
 Response Type          LogRT       
 Length:19392       Min.   : 2.811  
 Class :character   1st Qu.: 7.492  
 Mode  :character   Median : 8.107  
                    Mean   : 8.027  
                    3rd Qu.: 8.754  
                    Max.   :12.232  
```

#Stats

```r
HP_data <- data %>%
  mutate(
    Subject = as.factor(Subject),
    Condition = as.factor(Condition),
    Item = as.factor(Item)
    #Cond_Type = as.factor(Cond_Type),
    #Response = as.factor(Response)
  )
str(HP_data)
```

```
tibble [19,392 x 10] (S3: tbl_df/tbl/data.frame)
 $ Subject            : Factor w/ 404 levels "9898448","9898449",..: 13 13 13 13 13 13 13 1
 $ Item               : Factor w/ 48 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Tree Node Key      : chr [1:19392] "task-slet" "task-slet" "task-slet" "task-slet" ...
 $ RT                 : num [1:19392] 10352 4506 6214 5762 7446 ...
 $ Accuracy           : num [1:19392] 1 1 1 1 1 1 1 1 1 1 ...
 $ Spreadsheet: display: chr [1:19392] "Testing02" "Testing02" "Testing02" "Testing02" ...
 $ Screen             : chr [1:19392] "Testing" "Testing" "Testing" "Testing" ...
 $ Condition          : Factor w/ 6 levels "1e-1i","1e-2",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Response Type      : chr [1:19392] "response" "response" "response" "response" ...
 $ LogRT              : num [1:19392] 9.24 8.41 8.73 8.66 8.92 ...
```

```r
HP_data$Condition <-relevel(HP_data$Condition, ref="1e-1i")
```

```r
RT_model_1 <- lmer(LogRT ~ Condition + (1|Subject) + (1|Item), data = HP_data, REML = F)
summary(RT_model_1)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: LogRT ~ Condition + (1 | Subject) + (1 | Item)
   Data: HP_data

     AIC      BIC   logLik deviance df.resid
 37339.6  37410.5 -18660.8  37321.6    19383

Scaled residuals:
    Min      1Q  Median      3Q     Max
-9.8912 -0.5815 -0.0555  0.5167  7.8253

Random effects:
 Groups   Name        Variance Std.Dev.
```

```
 Subject  (Intercept) 0.7329   0.8561
 Item     (Intercept) 0.1251   0.3537
 Residual             0.3602   0.6002
Number of obs: 19392, groups:  Subject, 404; Item, 48

Fixed effects:
                Estimate Std. Error         df t value Pr(>|t|)
(Intercept)     8.080737   0.121428 426.374017  66.547   <2e-16 ***
Condition1e-2   0.050243   0.154565 403.653064   0.325    0.745
Condition1e-3  -0.220706   0.147919 403.653066  -1.492    0.136
Condition1i-2   0.009701   0.154565 403.653064   0.063    0.950
Condition1i-3  -0.034886   0.151223 403.653066  -0.231    0.818
Condition2-3   -0.085942   0.149738 403.653067  -0.574    0.566
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) Condition1e-2 Condition1e-3 Condition1i-2 Condition1i-3
Condition1e-2 -0.647
Condition1e-3 -0.676  0.531
Condition1i-2 -0.647  0.508         0.531
Condition1i-3 -0.661  0.519         0.543         0.519
Conditin2-3   -0.668  0.524         0.548         0.524         0.536
```

```r
ACC_model_1 <- glmer(Accuracy ~ Condition + (1|Subject) + (1|Item), data = HP_data, family =
summary(ACC_model_1)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: Accuracy ~ Condition + (1 | Subject) + (1 | Item)
   Data: HP_data

     AIC      BIC   logLik deviance df.resid
 24420.9  24483.9 -12202.5  24404.9    19384

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2172 -1.0560  0.5304  0.7839  1.5866

Random effects:
 Groups  Name        Variance Std.Dev.
```

```
 Subject (Intercept) 0.459757 0.67805
 Item    (Intercept) 0.003554 0.05961
Number of obs: 19392, groups:  Subject, 404; Item, 48

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.821522   0.097558   8.421  < 2e-16 ***
Condition1e-2 -0.079669   0.135745  -0.587  0.55727
Condition1e-3 -0.379217   0.129404  -2.930  0.00338 **
Condition1i-2  0.005684   0.135773   0.042  0.96661
Condition1i-3 -0.328282   0.132309  -2.481  0.01310 *
Condition2-3  -0.302998   0.131022  -2.313  0.02075 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) Condition1e-2 Condition1e-3 Condition1i-2 Condition1i-3
Condition1e-2 -0.712
Condition1e-3 -0.748  0.536
Condition1i-2 -0.712  0.511         0.537
Condition1i-3 -0.731  0.524         0.551         0.524
Conditin2-3   -0.738  0.530         0.556         0.530         0.544
```

```r
ACC_model_2 <- glmer(Accuracy ~ Condition + (1|Subject) , data = HP_data, family = "binomial"
summary(ACC_model_2)
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: Accuracy ~ Condition + (1 | Subject)
   Data: HP_data

     AIC      BIC   logLik deviance df.resid
 24420.8  24475.9 -12203.4  24406.8    19385

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.2007 -1.0587  0.5328  0.7836  1.5544

Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.4589   0.6775
```

```
Number of obs: 19392, groups:  Subject, 404

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)    0.82088    0.09711   8.453  < 2e-16 ***
Condition1e-2 -0.07961    0.13564  -0.587  0.55728    
Condition1e-3 -0.37894    0.12929  -2.931  0.00338 ** 
Condition1i-2  0.00568    0.13568   0.042  0.96661    
Condition1i-3 -0.32802    0.13222  -2.481  0.01310 *  
Condition2-3  -0.30281    0.13093  -2.313  0.02074 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) Condition1e-2 Condition1e-3 Condition1i-2 Condition1i-3
Condition1e-2 -0.714                                                        
Condition1e-3 -0.750  0.536                                                 
Condition1i-2 -0.715  0.511         0.537                                   
Condition1i-3 -0.733  0.524         0.551         0.525                     
Conditin2-3   -0.741  0.530         0.556         0.530         0.544       
```

```r
anova(ACC_model_1,ACC_model_2)
```

```
Data: HP_data
Models:
ACC_model_2: Accuracy ~ Condition + (1 | Subject)
ACC_model_1: Accuracy ~ Condition + (1 | Subject) + (1 | Item)
            npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
ACC_model_2    7 24421 24476 -12203    24407                     
ACC_model_1    8 24421 24484 -12202    24405 1.8305  1     0.1761
```

0.1 Plot

```r
Data$Condition <- as.factor(Data$Condition)


# RT <- ggplot(Data, aes(x=Condition, y=LogRT)) +
#   geom_violin(aes(fill = Condition), trim = FALSE,show.legend = FALSE) +
#   ylab("Log RT") +
#   xlab("Conditions") +
```

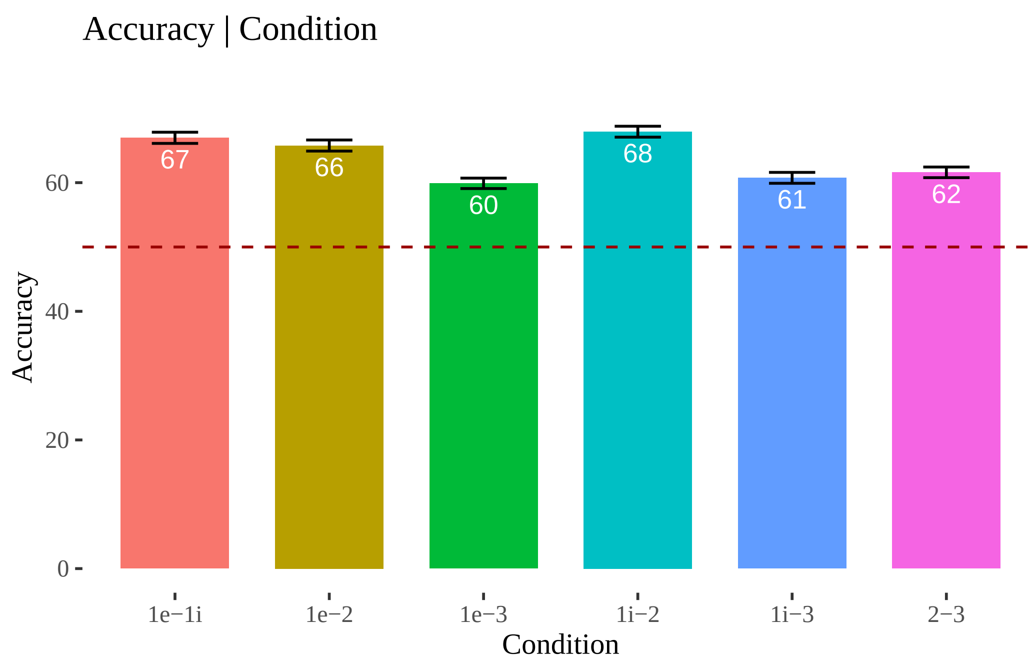
```
#
#   # geom_signif(
#   #   comparisons = list(c("Grammatical", "Pseudowords")),
#   #   margin_top = 0.20,
#   #   step_increase = 0.05,
#   #   tip_length = 0.01,
#   #   map_signif_level = TRUE
#   # )+
#   theme_Publication()+
#
#
#
#   # Add geom_boxplot() to include box plot
#   geom_boxplot(width = 0.2, fill = "white", color = "blue")
#
# RT + stat_summary(fun.data=mean_sdl, mult=1,
#                  geom="pointrange", color="red")
```

```
ACC_plot<- ggplot(data_group,
               aes(x=Condition,y=ACC,fill=Condition)) +

  # geom_bar function is used to plot bars of barplot
  geom_bar(stat = "identity", width = 0.7, position = position_dodge(0.7), show.legend = FALS
  #scale_x_discrete(limits = Conditions) +  facet_wrap( ~Prefix) +
  ylab("% Accept") +
  xlab("Conditions") +
  geom_text(aes(label=round(ACC)),vjust=1.6, color="white", size=3.5) +
  theme_Publication()+
  #theme_minimal() +
  #theme_clean() +
  #theme_wsj() +
  theme_tufte() +
  geom_hline(aes(yintercept=50), colour="#990000", linetype="dashed") +
  coord_cartesian(ylim = c(0, 75)) +
  # scale_y_continuous(expand = expansion(mult = c(0, 0.05)))+
  # geom_signif(
  #   comparisons = list(c("Grammatical", "Pseudowords")),
  #   margin_top = 0.12,
  #   step_increase = 0.09,
  #   tip_length = 0.05,
  #   annotation = c("***")
  # )+
```

```
#geom_errorbar function is used to plot error bars
geom_errorbar(aes(ymin=ACC-ACC_SE,
                  ymax=ACC+ACC_SE,
               width=0.3))
```

```
ACC_plot + labs(title="Accuracy | Condition",
        x ="Condition", y = "Accuracy")
```




```
report(ACC_model_1)
```

```
We fitted a logistic mixed model (estimated using ML and Nelder-Mead optimizer)
to predict Accuracy with Condition (formula: Accuracy ~ Condition). The model
included Subject as random effects (formula: list(~1 | Subject, ~1 | Item)).
The model's total explanatory power is weak (conditional R2 = 0.13) and the
part related to the fixed effects alone (marginal R2) is of 6.77e-03. The
model's intercept, corresponding to Condition = 1e-1i, is at 0.82 (95% CI
[0.63, 1.01], p < .001). Within this model:

  - The effect of Condition [1e-2] is statistically non-significant and negative
(beta = -0.08, 95% CI [-0.35, 0.19], p = 0.557; Std. beta = -0.08, 95% CI
[-0.35, 0.19])
```

```
  - The effect of Condition [1e-3] is statistically significant and negative
(beta = -0.38, 95% CI [-0.63, -0.13], p = 0.003; Std. beta = -0.38, 95% CI
[-0.63, -0.13])
  - The effect of Condition [1i-2] is statistically non-significant and positive
(beta = 5.68e-03, 95% CI [-0.26, 0.27], p = 0.967; Std. beta = 5.68e-03, 95% CI
[-0.26, 0.27])
  - The effect of Condition [1i-3] is statistically significant and negative
(beta = -0.33, 95% CI [-0.59, -0.07], p = 0.013; Std. beta = -0.33, 95% CI
[-0.59, -0.07])
  - The effect of Condition [2-3] is statistically significant and negative (beta
= -0.30, 95% CI [-0.56, -0.05], p = 0.021; Std. beta = -0.30, 95% CI [-0.56,
-0.05])

Standardized parameters were obtained by fitting the model on a standardized
version of the dataset. 95% Confidence Intervals (CIs) and p-values were
computed using a Wald z-distribution approximation.
```

```
BAN_ACC_Model1.emm <- emmeans(ACC_model_1, ~ Condition, lmer.df = "asymp")
pairs(BAN_ACC_Model1.emm, simple = "each")
```

```
 contrast          estimate    SE  df z.ratio p.value
 (1e-1i) - (1e-2)  0.07967 0.136 Inf   0.587  0.9919
 (1e-1i) - (1e-3)  0.37922 0.129 Inf   2.930  0.0397
 (1e-1i) - (1i-2) -0.00568 0.136 Inf  -0.042  1.0000
 (1e-1i) - (1i-3)  0.32828 0.132 Inf   2.481  0.1296
 (1e-1i) - (2-3)   0.30300 0.131 Inf   2.313  0.1888
 (1e-2) - (1e-3)   0.29955 0.128 Inf   2.344  0.1764
 (1e-2) - (1i-2)  -0.08535 0.134 Inf  -0.636  0.9883
 (1e-2) - (1i-3)   0.24861 0.131 Inf   1.901  0.4011
 (1e-2) - (2-3)    0.22333 0.129 Inf   1.726  0.5147
 (1e-3) - (1i-2)  -0.38490 0.128 Inf  -3.012  0.0311
 (1e-3) - (1i-3)  -0.05094 0.124 Inf  -0.410  0.9985
 (1e-3) - (2-3)   -0.07622 0.123 Inf  -0.621  0.9895
 (1i-2) - (1i-3)   0.33397 0.131 Inf   2.554  0.1089
 (1i-2) - (2-3)    0.30868 0.129 Inf   2.385  0.1614
 (1i-3) - (2-3)   -0.02528 0.126 Inf  -0.201  1.0000

Results are given on the log odds ratio (not the response) scale.
P value adjustment: tukey method for comparing a family of 6 estimates
```