Numbers-Experment-MML

Behavioural analysis

Swarnendu Moitra

Table of contents

0.1 Plot 6

Analysis based on Persons experiment: Middleton and Moitra 2023

```
library("tidyverse")
library("plotrix")
library("lme4")
library("lmerTest")
library("report")
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 %>%
        select(`Participant Private ID`,`Trial Number`,'Tree Node Key', 'Reaction Time', Cor
        rename(Subject = `Participant Private ID`,
                      Item = `Trial Number`,
                      Condition = 'Tree Node Key',
                      RT = 'Reaction Time',
                      Accuracy = Correct) %>%
        filter(Screen == "Testing",
                `Response Type` == "response") %>%
        mutate(LogRT = log(RT))
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)
```

```r
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        Condition               RT         
 Min.   : 9898448   Min.   : 1.00   Length:18096       Min.   :    16.63  
 1st Qu.: 9992389   1st Qu.:12.75   Class :character   1st Qu.:  1757.67  
 Median : 9996949   Median :24.50   Mode  :character   Median :  3311.54  
 Mean   :10339737   Mean   :24.50                      Mean   :  5327.17  
 3rd Qu.:10840811   3rd Qu.:36.25                      3rd Qu.:  6339.98  
 Max.   :10854219   Max.   :48.00                      Max.   :205295.10  
    Accuracy      Spreadsheet: display    Screen           Task Name        
 Min.   :0.0000   Length:18096         Length:18096       Length:18096      
 1st Qu.:0.0000   Class :character     Class :character   Class :character  
 Median :1.0000   Mode  :character     Mode  :character   Mode  :character  
 Mean   :0.6348                                                             
 3rd Qu.:1.0000                                                             
 Max.   :1.0000                                                             
 Response Type          LogRT       
 Length:18096       Min.   : 2.811  
 Class :character   1st Qu.: 7.472  
 Mode  :character   Median : 8.105  
                    Mean   : 8.019  
                    3rd Qu.: 8.755  
                    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),
```

```r
    #Response = as.factor(Response)
  )
str(HP_data)
```

```
tibble [18,096 x 10] (S3: tbl_df/tbl/data.frame)
 $ Subject             : Factor w/ 377 levels "9898448","9898449",..: 13 13 13 13 13 13 13 13
 $ Item                : Factor w/ 48 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Condition           : Factor w/ 6 levels "task-98zo","task-ajpl",..: 6 6 6 6 6 6 6 6 6 6
 $ RT                  : num [1:18096] 10352 4506 6214 5762 7446 ...
 $ Accuracy            : num [1:18096] 1 1 1 1 1 1 1 1 1 1 ...
 $ Spreadsheet: display: chr [1:18096] "Testing02" "Testing02" "Testing02" "Testing02" ...
 $ Screen              : chr [1:18096] "Testing" "Testing" "Testing" "Testing" ...
 $ Task Name           : chr [1:18096] "1e-1i B Testing" "1e-1i B Testing" "1e-1i B Testing"
 $ Response Type       : chr [1:18096] "response" "response" "response" "response" ...
 $ LogRT               : num [1:18096] 9.24 8.41 8.73 8.66 8.92 ...
```

```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
 35035.5  35105.7 -17508.7  35017.5    18087

Scaled residuals:
    Min      1Q  Median      3Q     Max
-9.8280 -0.5793 -0.0551  0.5167  7.7930

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.7502   0.8661
 Item     (Intercept) 0.1259   0.3548
 Residual             0.3637   0.6031
Number of obs: 18096, groups:  Subject, 377; Item, 48

Fixed effects:
                     Estimate Std. Error         df t value Pr(>|t|)
(Intercept)          8.101864   0.125249 410.069412  64.686   <2e-16 ***
```

```
Conditiontask-ajpl  -0.032884   0.156126 376.676246  -0.211   0.8333
Conditiontask-goty  -0.266378   0.154566 376.676246  -1.723   0.0856 .
Conditiontask-hebw  -0.007211   0.160296 376.676253  -0.045   0.9641
Conditiontask-ld3j  -0.106709   0.156674 376.676247  -0.681   0.4962
Conditiontask-slet  -0.052394   0.163086 376.676258  -0.321   0.7482
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Cndtntsk-j Cndtntsk-g Cndtntsk-h Cndt-3
Cndtntsk-jp -0.668
Cndtntsk-gt -0.675  0.541
Cndtntsk-hb -0.651  0.522      0.527
Cndtntsk-l3 -0.666  0.534      0.540      0.520
Cndtntsk-sl -0.640  0.513      0.518      0.500      0.511
```

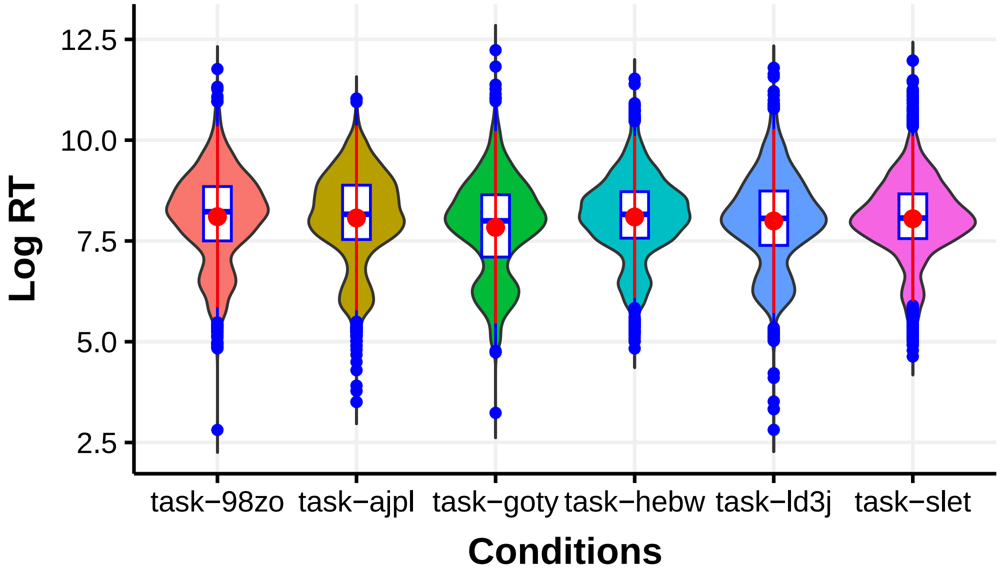
```r
ACC_model_1 <- glmer(Accuracy ~ Condition + (1|Subject) + (1|Item), data = HP_data, family =
summary(ACC_model_1)
```

```
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
 22852.5  22914.9 -11418.2  22836.5    18088

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1337 -1.0617  0.5341  0.7855  1.5824

Random effects:
 Groups  Name        Variance Std.Dev.
 Subject (Intercept) 0.445989 0.66782
 Item    (Intercept) 0.003922 0.06263
Number of obs: 18096, groups:  Subject, 377; Item, 48

Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         0.68253    0.09795   6.968 3.22e-12 ***
Conditiontask-ajpl -0.19239    0.13274  -1.449   0.1472
```

```
Conditiontask-goty -0.21791    0.13132  -1.659   0.0971 .
Conditiontask-hebw  0.15431    0.13688   1.127   0.2596
Conditiontask-ld3j -0.13310    0.13323  -0.999   0.3178
Conditiontask-slet  0.07666    0.13938   0.550   0.5823
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) Cndtntsk-j Cndtntsk-g Cndtntsk-h Cndt-3
Cndtntsk-jp -0.731
Cndtntsk-gt -0.739  0.545
Cndtntsk-hb -0.709  0.523      0.528
Cndtntsk-l3 -0.728  0.537      0.543      0.521
Cndtntsk-sl -0.696  0.513      0.519      0.498      0.512
```

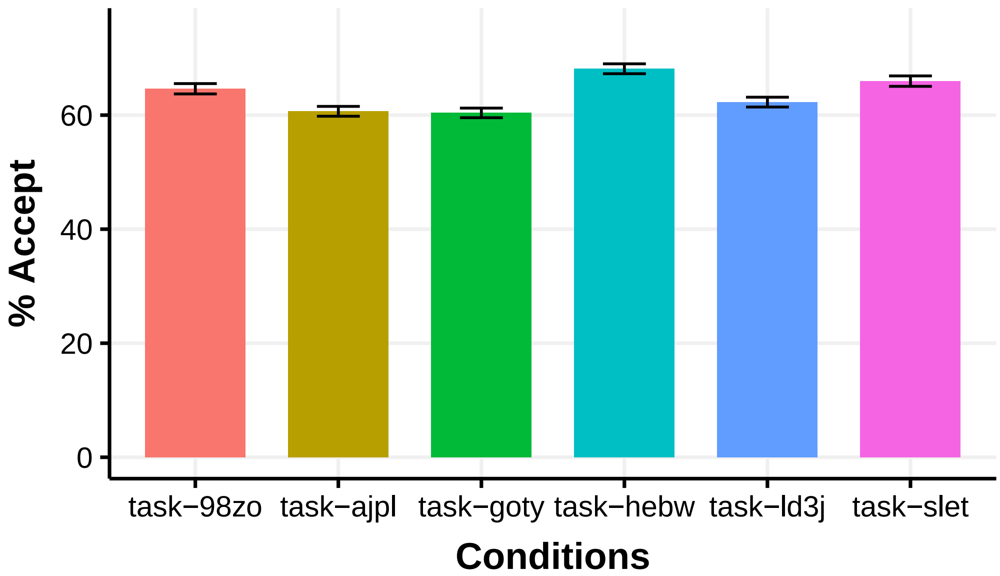
0.1 Plot

```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") +
  theme_Publication()+
  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))
```

```r
ACC_plot
```




```r
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 = 0.12) and the
part related to the fixed effects alone (marginal R2) is of 5.14e-03. The
model's intercept, corresponding to Condition = task-98zo, is at 0.68 (95% CI
[0.49, 0.87], p < .001). Within this model:

  - The effect of Condition [task-ajpl] is statistically non-significant and
negative (beta = -0.19, 95% CI [-0.45, 0.07], p = 0.147; Std. beta = -0.19, 95%
CI [-0.45, 0.07])
  - The effect of Condition [task-goty] is statistically non-significant and
negative (beta = -0.22, 95% CI [-0.48, 0.04], p = 0.097; Std. beta = -0.22, 95%
CI [-0.48, 0.04])
  - The effect of Condition [task-hebw] is statistically non-significant and
positive (beta = 0.15, 95% CI [-0.11, 0.42], p = 0.260; Std. beta = 0.15, 95%
CI [-0.11, 0.42])
  - The effect of Condition [task-ld3j] is statistically non-significant and
```

negative (beta = -0.13, 95% CI [-0.39, 0.13], p = 0.318; Std. beta = -0.13, 95% CI [-0.39, 0.13])
  - The effect of Condition [task-slet] is statistically non-significant and positive (beta = 0.08, 95% CI [-0.20, 0.35], p = 0.582; Std. beta = 0.08, 95% CI [-0.20, 0.35])

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.