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")
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")
```

```r
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:10608       Min.   :    16.63  
 1st Qu.: 9991270   1st Qu.:12.75   Class :character   1st Qu.:  1749.92  
 Median : 9992609   Median :24.50   Mode  :character   Median :  3241.49  
 Mean   : 9982939   Mean   :24.50                      Mean   :  5158.65  
 3rd Qu.: 9994705   3rd Qu.:36.25                      3rd Qu.:  5978.85  
 Max.   :10006213   Max.   :48.00                      Max.   :158673.60  
    Accuracy     Spreadsheet: display    Screen            Task Name        
 Min.   :0.000   Length:10608         Length:10608       Length:10608      
 1st Qu.:0.000   Class :character     Class :character   Class :character  
 Median :1.000   Mode  :character     Mode  :character   Mode  :character  
 Mean   :0.629                                                             
 3rd Qu.:1.000                                                             
 Max.   :1.000                                                             
 Response Type          LogRT       
 Length:10608       Min.   : 2.811  
 Class :character   1st Qu.: 7.467  
 Mode  :character   Median : 8.084  
                    Mean   : 7.986  
                    3rd Qu.: 8.696  
                    Max.   :11.975  
```

#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 [10,608 x 10] (S3: tbl_df/tbl/data.frame)
 $ Subject             : Factor w/ 221 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 [1:10608] 10352 4506 6214 5762 7446 ...
 $ Accuracy            : num [1:10608] 1 1 1 1 1 1 1 1 1 1 ...
 $ Spreadsheet: display: chr [1:10608] "Testing02" "Testing02" "Testing02" "Testing02" ...
 $ Screen              : chr [1:10608] "Testing" "Testing" "Testing" "Testing" ...
 $ Task Name           : chr [1:10608] "1e-1i B Testing" "1e-1i B Testing" "1e-1i B Testing"
 $ Response Type       : chr [1:10608] "response" "response" "response" "response" ...
 $ LogRT               : num [1:10608] 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
 20732.2  20797.6 -10357.1  20714.2    10599

Scaled residuals:
    Min      1Q  Median      3Q     Max
-9.6906 -0.5697 -0.0638  0.5104  7.3289

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 0.7382   0.8592
 Item     (Intercept) 0.1277   0.3573
 Residual             0.3679   0.6066
Number of obs: 10608, groups:  Subject, 221; Item, 48

Fixed effects:
                      Estimate Std. Error       df t value Pr(>|t|)
(Intercept)             8.1802     0.1712 254.7636  47.790   <2e-16 ***
```

```
Conditiontask-ajpl  -0.1761     0.2151 220.8044  -0.819   0.4139
Conditiontask-goty  -0.3840     0.2079 220.8044  -1.847   0.0661 .
Conditiontask-hebw  -0.1464     0.2151 220.8044  -0.680   0.4969
Conditiontask-ld3j  -0.1779     0.2151 220.8044  -0.827   0.4091
Conditiontask-slet  -0.1946     0.2204 220.8044  -0.883   0.3783
---
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 -0.724
Cndtntsk-gt -0.749  0.596
Cndtntsk-hb -0.724  0.576      0.596
Cndtntsk-l3 -0.724  0.576      0.596      0.576
Cndtntsk-sl -0.706  0.562      0.581      0.562      0.562
```

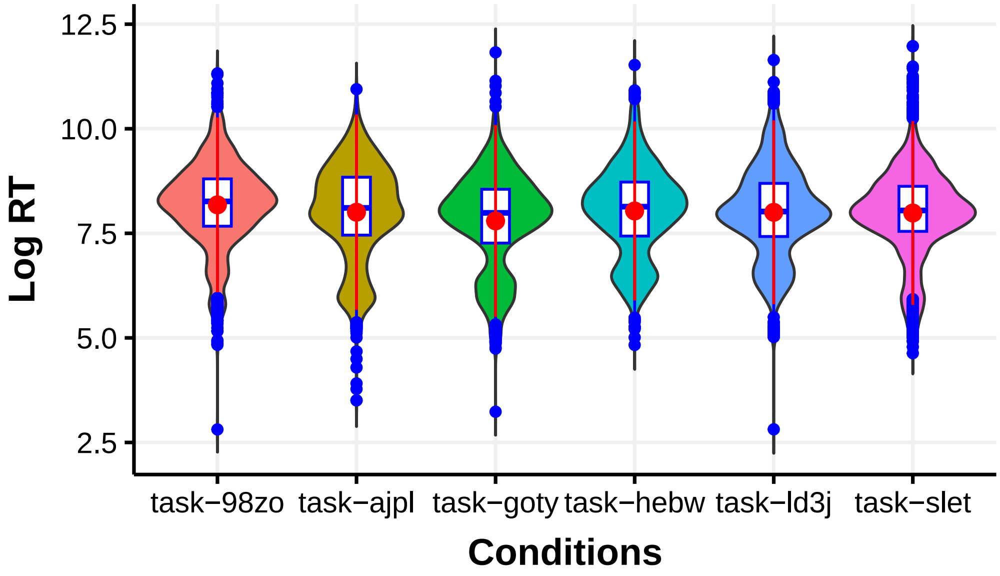
```r
ACC_model_1 <- glmer(Accuracy ~ Condition + (1|Subject) + (1|Item), data = HP_data, family =
summary(ACC_model_1)
```

```
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
 13403.1  13461.3  -6693.6  13387.1    10600

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1499 -1.0355  0.5338  0.7934  1.5746

Random effects:
 Groups  Name        Variance  Std.Dev.
 Subject (Intercept) 0.4869742 0.69784
 Item    (Intercept) 0.0009203 0.03034
Number of obs: 10608, groups:  Subject, 221; Item, 48

Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.6338     0.1453   4.361 1.29e-05 ***
Conditiontask-ajpl  -0.2003     0.1911  -1.048    0.295
```

```
Conditiontask-goty  -0.1563     0.1846  -0.847    0.397
Conditiontask-hebw   0.2270     0.1918   1.183    0.237
Conditiontask-ld3j  -0.1341     0.1911  -0.702    0.483
Conditiontask-slet   0.1027     0.1964   0.523    0.601
---
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 -0.759
Cndtntsk-gt -0.786  0.598
Cndtntsk-hb -0.756  0.575      0.595
Cndtntsk-l3 -0.759  0.577      0.598      0.575
Cndtntsk-sl -0.738  0.562      0.581      0.560      0.562
```

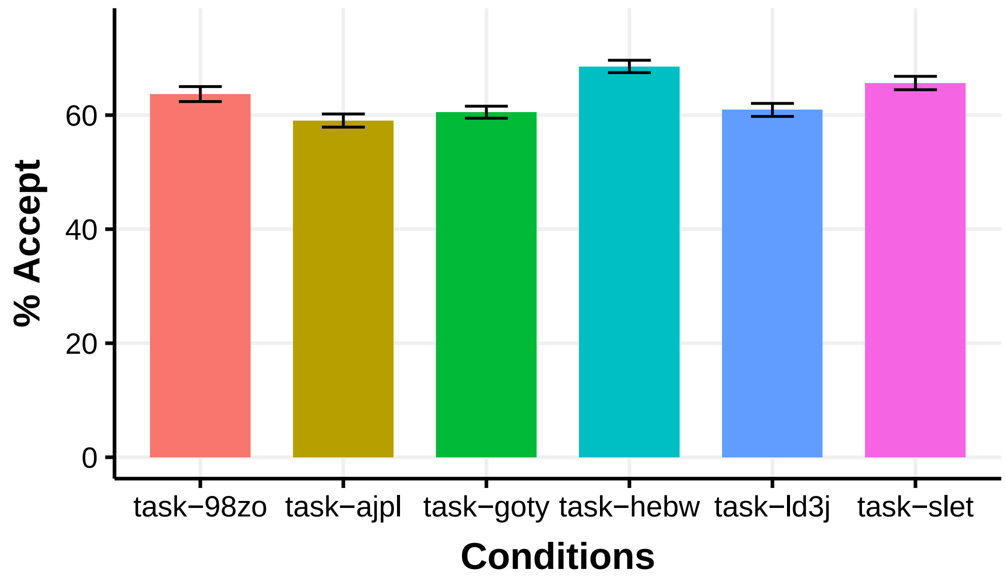
0.1 Plot

```
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))
```

```
ACC_plot
```




```
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 moderate (conditional R2 = 0.13) and the
part related to the fixed effects alone (marginal R2) is of 6.38e-03. The
model's intercept, corresponding to Condition = task-98zo, is at 0.63 (95% CI
[0.35, 0.92], p < .001). Within this model:

  - The effect of Condition [task-ajpl] is statistically non-significant and
negative (beta = -0.20, 95% CI [-0.57, 0.17], p = 0.295; Std. beta = -0.20, 95%
CI [-0.57, 0.17])
  - The effect of Condition [task-goty] is statistically non-significant and
negative (beta = -0.16, 95% CI [-0.52, 0.21], p = 0.397; Std. beta = -0.16, 95%
CI [-0.52, 0.21])
  - The effect of Condition [task-hebw] is statistically non-significant and
positive (beta = 0.23, 95% CI [-0.15, 0.60], p = 0.237; Std. beta = 0.23, 95%
CI [-0.15, 0.60])
  - The effect of Condition [task-ld3j] is statistically non-significant and
```

negative (beta = -0.13, 95% CI [-0.51, 0.24], p = 0.483; Std. beta = -0.13, 95% CI [-0.51, 0.24])

- The effect of Condition [task-slet] is statistically non-significant and positive (beta = 0.10, 95% CI [-0.28, 0.49], p = 0.601; Std. beta = 0.10, 95% CI [-0.28, 0.49])

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.