statistical analysis

Mingjia

3/8/2022

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

library(ggplot2)  
library(dplyr)  
library(purrr)

##   
## Attaching package: 'purrr'

## The following object is masked from 'package:data.table':  
##   
## transpose

## The following object is masked from 'package:car':  
##   
## some

library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

source("utils.R")  
# ---- Load the Data ----  
data\_subj = fread("summary data/test\_accuracy.csv")

#------------------median split IDs-----------------------  
data\_subj\_4cat = filter(data\_subj, condition == "4cat")  
  
below\_4cat\_idx = with(data\_subj\_4cat,order(train[condition == "4cat"]))[1:30]  
below\_4cat\_ID = with(data\_subj\_4cat,ID[below\_4cat\_idx])  
data\_subj\_4cat$medgrp = ifelse(data\_subj\_4cat$ID %in% below\_4cat\_ID,"below","above")  
  
data\_subj\_2cat = filter(data\_subj, condition == "2cat")  
  
below\_2cat\_idx = with(data\_subj\_2cat,order(train[condition == "2cat"]))[1:30]  
below\_2cat\_ID = with(data\_subj\_2cat,ID[below\_2cat\_idx])  
data\_subj\_2cat$medgrp = ifelse(data\_subj\_2cat$ID %in% below\_2cat\_ID,"below","above")  
  
data\_subj = rbind(data\_subj\_2cat,data\_subj\_4cat)

# Accuracy on training items

## ANOVA: conditions x accuracy group

res.aov <- anova\_test(data=data\_subj,train ~ condition \* medgrp,   
 type=3, effect.size = "pes",detailed=T)

## Coefficient covariances computed by hccm()

get\_anova\_table(res.aov,correction = "none")

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F p p<.05 pes  
## 1 (Intercept) 74.207 1.297 1 117 6695.336 4.09e-105 \* 0.983000  
## 2 condition 0.001 1.297 1 117 0.105 7.46e-01 0.000899  
## 3 medgrp 1.868 1.297 1 117 168.574 2.03e-24 \* 0.590000  
## 4 condition:medgrp 0.121 1.297 1 117 10.953 1.00e-03 \* 0.086000

Prcorr\_stats = data\_subj %>%   
 group\_by(condition,medgrp) %>%   
 summarize(mean = mean(train),  
 SD = sd(train),  
 SEM = SD/sqrt(length(train)))

## `summarise()` has grouped output by 'condition'. You can override using the `.groups` argument.

show(Prcorr\_stats)

## # A tibble: 4 x 5  
## # Groups: condition [2]  
## condition medgrp mean SD SEM  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 2cat above 0.873 0.0702 0.0126   
## 2 2cat below 0.688 0.0868 0.0159   
## 3 4cat above 0.942 0.0498 0.00909  
## 4 4cat below 0.630 0.172 0.0314

The main effect of the median split groups is significant, so was the interaction between the condition and the median split groups.

## t-test: compare 2cat and 4cat conditions for above-median group

Prcorr\_above = filter(data\_subj,medgrp == "above")  
  
t.test(data = Prcorr\_above, train ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: train by condition  
## t = -4.4517, df = 59, p-value = 3.847e-05  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.10083534 -0.03829676  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.8726959 0.9422619

Prcorr\_above\_stats = Prcorr\_above %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(train),  
 SEM = sd(train)/sqrt(length(train)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.873 0.0126   
## 2 4cat 0.942 0.00909

For the above-median groups, the training-item accuracy in the 4-cat condition is significantly higher than in the 2-cat condition.

## t-test: compare 2cat and 4cat conditions for below-median group

Prcorr\_below = filter(data\_subj,medgrp == "below")  
  
t.test(data = Prcorr\_below, train ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: train by condition  
## t = 1.624, df = 58, p-value = 0.1098  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.01328948 0.12757519  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.6875000 0.6303571

Prcorr\_above\_stats = Prcorr\_below %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(train),  
 SEM = sd(train)/sqrt(length(train)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.688 0.0159  
## 2 4cat 0.630 0.0314

For the below-median groups, the training-item accuracy is not significantly different between the two conditions.

# Accuracy on critical items

## ANOVA: conditions x accuracy group

res.aov <- anova\_test(data=data\_subj,critical ~ condition \* medgrp,   
 type=3, effect.size = "pes",detailed=T)

## Coefficient covariances computed by hccm()

get\_anova\_table(res.aov,correction = "none")

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F p p<.05 pes  
## 1 (Intercept) 22.441 10.426 1 117 251.825 6.00e-31 \* 0.683  
## 2 condition 1.148 10.426 1 117 12.887 4.85e-04 \* 0.099  
## 3 medgrp 0.808 10.426 1 117 9.070 3.00e-03 \* 0.072  
## 4 condition:medgrp 1.173 10.426 1 117 13.164 4.24e-04 \* 0.101

Prcorr\_stats = data\_subj %>%   
 group\_by(condition,medgrp) %>%   
 summarize(mean = mean(critical),  
 SD = sd(critical),  
 SEM = SD/sqrt(length(critical)))

## `summarise()` has grouped output by 'condition'. You can override using the `.groups` argument.

show(Prcorr\_stats)

## # A tibble: 4 x 5  
## # Groups: condition [2]  
## condition medgrp mean SD SEM  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 2cat above 0.317 0.343 0.0616  
## 2 2cat below 0.35 0.183 0.0334  
## 3 4cat above 0.708 0.365 0.0667  
## 4 4cat below 0.348 0.266 0.0486

The main effects of condition and median split are both significant on the accuracy of critical items, so was the interaction between the condition and the median split groups.

## t-test: compare 2cat and 4cat conditions for above-median group

Prcorr\_above = filter(data\_subj,medgrp == "above")  
  
t.test(data = Prcorr\_above, critical ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: critical by condition  
## t = -4.3195, df = 59, p-value = 6.077e-05  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.5733005 -0.2103017  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.3165323 0.7083333

Prcorr\_above\_stats = Prcorr\_above %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(critical),  
 SEM = sd(critical)/sqrt(length(critical)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.317 0.0616  
## 2 4cat 0.708 0.0667

For the above-median groups, the accuracy on the critical items in the 4-cat condition is significantly higher than in the 2-cat condition.

## t-test: compare 2cat and 4cat conditions for below-median group

Prcorr\_below = filter(data\_subj,medgrp == "below")  
  
t.test(data = Prcorr\_below, critical ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: critical by condition  
## t = 0.035327, df = 58, p-value = 0.9719  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.1159630 0.1201297  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.3500000 0.3479167

Prcorr\_above\_stats = Prcorr\_below %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(critical),  
 SEM = sd(critical)/sqrt(length(critical)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.35 0.0334  
## 2 4cat 0.348 0.0486

For the below-median groups, the accuracy on the critical items is not significantly different between the two conditions.

# 

# Accuracy on other transfer items

## ANOVA: conditions x accuracy group

res.aov <- anova\_test(data=data\_subj,other ~ condition \* medgrp,   
 type=3, effect.size = "pes",detailed=T)

## Coefficient covariances computed by hccm()

get\_anova\_table(res.aov,correction = "none")

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F p p<.05 pes  
## 1 (Intercept) 67.820 2.137 1 117 3712.959 1.76e-90 \* 0.969  
## 2 condition 0.080 2.137 1 117 4.388 3.80e-02 \* 0.036  
## 3 medgrp 1.012 2.137 1 117 55.387 1.82e-11 \* 0.321  
## 4 condition:medgrp 0.177 2.137 1 117 9.684 2.00e-03 \* 0.076

Prcorr\_stats = data\_subj %>%   
 group\_by(condition,medgrp) %>%   
 summarize(mean = mean(critical),  
 SD = sd(critical),  
 SEM = SD/sqrt(length(critical)))

## `summarise()` has grouped output by 'condition'. You can override using the `.groups` argument.

show(Prcorr\_stats)

## # A tibble: 4 x 5  
## # Groups: condition [2]  
## condition medgrp mean SD SEM  
## <chr> <chr> <dbl> <dbl> <dbl>  
## 1 2cat above 0.317 0.343 0.0616  
## 2 2cat below 0.35 0.183 0.0334  
## 3 4cat above 0.708 0.365 0.0667  
## 4 4cat below 0.348 0.266 0.0486

The main effects of condition and median split are both significant on the accuracy of the other transfer items, so was the interaction between the condition and the median split groups.

## t-test: compare 2cat and 4cat conditions for above-median group

Prcorr\_above = filter(data\_subj,medgrp == "above")  
  
t.test(data = Prcorr\_above, other ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: other by condition  
## t = -4.1778, df = 59, p-value = 9.859e-05  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.18924359 -0.06667039  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.7762097 0.9041667

Prcorr\_above\_stats = Prcorr\_above %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(other),  
 SEM = sd(other)/sqrt(length(other)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.776 0.0221  
## 2 4cat 0.904 0.0211

For the above-median groups, the accuracy on the other transfer items in the 4-cat condition is significantly higher than in the 2-cat condition.

## t-test: compare 2cat and 4cat conditions for below-median group

Prcorr\_below = filter(data\_subj,medgrp == "below")  
  
t.test(data = Prcorr\_below, other ~ condition, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: other by condition  
## t = 0.64841, df = 58, p-value = 0.5193  
## alternative hypothesis: true difference in means between group 2cat and group 4cat is not equal to 0  
## 95 percent confidence interval:  
## -0.0521777 0.1021777  
## sample estimates:  
## mean in group 2cat mean in group 4cat   
## 0.6697917 0.6447917

Prcorr\_above\_stats = Prcorr\_below %>%   
 group\_by(condition) %>%   
 summarize(mean = mean(other),  
 SEM = sd(other)/sqrt(length(other)))  
show(Prcorr\_above\_stats)

## # A tibble: 2 x 3  
## condition mean SEM  
## <chr> <dbl> <dbl>  
## 1 2cat 0.670 0.0212  
## 2 4cat 0.645 0.0322

For the below-median groups, the accuracy on the other transfer items is not significantly different between the two conditions.