lmer\_w2v.R

linwang

2023-08-29

rm(list=ls())  
library(vioplot)

## Loading required package: sm

## Package 'sm', version 2.2-5.7: type help(sm) for summary information

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

library(emmeans)  
library(lme4)

## Loading required package: Matrix

library(lmerTest)

##   
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':  
##   
## lmer

## The following object is masked from 'package:stats':  
##   
## step

library(tidyr)

##   
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':  
##   
## expand, pack, unpack

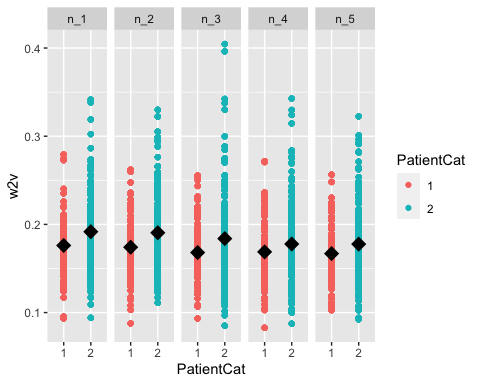
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

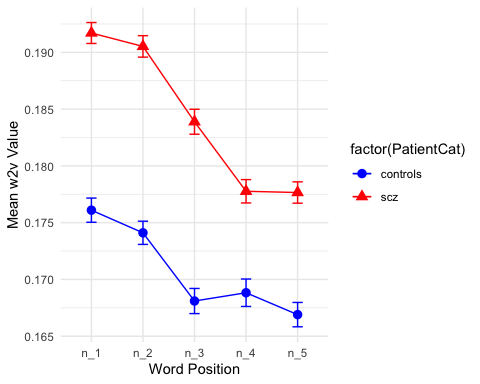
library(ggplot2)  
library(hypr)  
  
setwd("/Users/linwang/Dropbox (Partners HealthCare)/OngoingProjects/sczTopic/stimuli/")  
  
############################prepare data###################  
data <- read.csv(file = 'TOPSY\_subjectspec\_variables.csv')  
  
df <- data[,c('ID','PatientCat','Gender','AgeScan1','stim','TLI\_DISORG','n\_sentence','stim','n\_1','n\_2','n\_3','n\_4','n\_5')]  
df <- subset(df, PatientCat == 1 | PatientCat == 2)  
  
df2 <- df %>%  
 pivot\_longer(cols = c(n\_1, n\_2, n\_3, n\_4, n\_5), names\_to = "wordpos", values\_to = "w2v")  
  
df2$wordpos <- relevel(factor(df2$wordpos), ref = 'n\_1')  
df2$PatientCat <- relevel(factor(df2$PatientCat), ref = '1')  
  
# Plot  
df2 %>%   
 ggplot(aes(x = PatientCat, y = w2v, col = PatientCat)) +  
 geom\_point() +   
 facet\_grid(cols = vars(wordpos)) +   
 stat\_summary(fun = mean, na.rm = TRUE,   
 geom = "point", shape = "diamond",   
 color = "black", size = 5)



# get mean and standard error values  
summary\_data <- df2 %>%  
 group\_by(wordpos, PatientCat) %>%  
 summarise(mean\_w2v = mean(w2v, na.rm = TRUE),  
 se\_w2v = sd(w2v, na.rm = TRUE) / sqrt(n()))

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

# Create the line plot with standard error bars  
ggplot(summary\_data, aes(x = wordpos, y = mean\_w2v, color = factor(PatientCat), group = PatientCat)) +  
 geom\_line() +  
 geom\_point(aes(shape = factor(PatientCat)), size = 3) +  
 geom\_errorbar(aes(ymin = mean\_w2v - se\_w2v, ymax = mean\_w2v + se\_w2v), width = 0.2) +  
 labs(x = "Word Position", y = "Mean w2v Value") +  
 scale\_color\_manual(values = c("blue", "red"), labels = c("controls", "scz")) +  
 scale\_shape\_manual(values = c(16, 17), labels = c("controls", "scz")) +  
 theme\_minimal()



############################run models###################  
  
mpos1 <- lmer(w2v ~ PatientCat\*wordpos + TLI\_DISORG + n\_sentence + (1 | ID), data = df2)  
anova(mpos1)

## Type III Analysis of Variance Table with Satterthwaite's method  
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)   
## PatientCat 0.004345 0.004345 1 99.9 3.6919 0.0575280 .   
## wordpos 0.300973 0.075243 4 13947.9 63.9293 < 2.2e-16 \*\*\*  
## TLI\_DISORG 0.002918 0.002918 1 100.1 2.4788 0.1185444   
## n\_sentence 0.000670 0.000670 1 100.9 0.5696 0.4521595   
## PatientCat:wordpos 0.021916 0.005479 4 13947.9 4.6551 0.0009377 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(mpos1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: w2v ~ PatientCat \* wordpos + TLI\_DISORG + n\_sentence + (1 | ID)  
## Data: df2  
##   
## REML criterion at convergence: -54390.3  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.7488 -0.6386 -0.0966 0.5956 4.5625   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.0004863 0.02205   
## Residual 0.0011770 0.03431   
## Number of obs: 14060, groups: ID, 104  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.814e-01 7.035e-03 1.045e+02 25.786 < 2e-16 \*\*\*  
## PatientCat2 1.069e-02 5.216e-03 1.123e+02 2.049 0.0428 \*   
## wordposn\_2 -3.437e-03 1.618e-03 1.395e+04 -2.124 0.0337 \*   
## wordposn\_3 -8.832e-03 1.618e-03 1.395e+04 -5.458 4.89e-08 \*\*\*  
## wordposn\_4 -8.938e-03 1.618e-03 1.395e+04 -5.524 3.38e-08 \*\*\*  
## wordposn\_5 -1.101e-02 1.618e-03 1.395e+04 -6.805 1.05e-11 \*\*\*  
## TLI\_DISORG 3.786e-03 2.405e-03 1.001e+02 1.574 0.1185   
## n\_sentence -3.739e-04 4.954e-04 1.009e+02 -0.755 0.4522   
## PatientCat2:wordposn\_2 2.256e-03 1.962e-03 1.395e+04 1.150 0.2502   
## PatientCat2:wordposn\_3 1.015e-03 1.962e-03 1.395e+04 0.517 0.6050   
## PatientCat2:wordposn\_4 -5.004e-03 1.962e-03 1.395e+04 -2.550 0.0108 \*   
## PatientCat2:wordposn\_5 -3.038e-03 1.962e-03 1.395e+04 -1.549 0.1215   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) PtntC2 wrdp\_2 wrdp\_3 wrdp\_4 wrdp\_5 TLI\_DI n\_sntn PC2:\_2  
## PatientCat2 -0.533   
## wordposn\_2 -0.115 0.155   
## wordposn\_3 -0.115 0.155 0.500   
## wordposn\_4 -0.115 0.155 0.500 0.500   
## wordposn\_5 -0.115 0.155 0.500 0.500 0.500   
## TLI\_DISORG 0.298 -0.384 0.000 0.000 0.000 0.000   
## n\_sentence -0.826 0.152 0.000 0.000 0.000 0.000 -0.417   
## PtntCt2:w\_2 0.095 -0.188 -0.825 -0.412 -0.412 -0.412 0.000 0.000   
## PtntCt2:w\_3 0.095 -0.188 -0.412 -0.825 -0.412 -0.412 0.000 0.000 0.500  
## PtntCt2:w\_4 0.095 -0.188 -0.412 -0.412 -0.825 -0.412 0.000 0.000 0.500  
## PtntCt2:w\_5 0.095 -0.188 -0.412 -0.412 -0.412 -0.825 0.000 0.000 0.500  
## PC2:\_3 PC2:\_4  
## PatientCat2   
## wordposn\_2   
## wordposn\_3   
## wordposn\_4   
## wordposn\_5   
## TLI\_DISORG   
## n\_sentence   
## PtntCt2:w\_2   
## PtntCt2:w\_3   
## PtntCt2:w\_4 0.500   
## PtntCt2:w\_5 0.500 0.500

mpos2 = lmer(w2v ~ PatientCat\*wordpos + (1 | ID), data = df2)  
summary(mpos2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: w2v ~ PatientCat \* wordpos + (1 | ID)  
## Data: df2  
##   
## REML criterion at convergence: -55491.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.7532 -0.6359 -0.1011 0.5915 4.5683   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.0004796 0.02190   
## Residual 0.0011744 0.03427   
## Number of obs: 14330, groups: ID, 106  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.763e-01 3.817e-03 1.191e+02 46.193 < 2e-16 \*\*\*  
## PatientCat2 1.505e-02 4.696e-03 1.190e+02 3.204 0.001740 \*\*   
## wordposn\_2 -1.994e-03 1.570e-03 1.422e+04 -1.270 0.204106   
## wordposn\_3 -8.001e-03 1.570e-03 1.422e+04 -5.097 3.50e-07 \*\*\*  
## wordposn\_4 -7.272e-03 1.570e-03 1.422e+04 -4.632 3.65e-06 \*\*\*  
## wordposn\_5 -9.200e-03 1.570e-03 1.422e+04 -5.860 4.73e-09 \*\*\*  
## PatientCat2:wordposn\_2 8.128e-04 1.922e-03 1.422e+04 0.423 0.672306   
## PatientCat2:wordposn\_3 1.843e-04 1.922e-03 1.422e+04 0.096 0.923606   
## PatientCat2:wordposn\_4 -6.669e-03 1.922e-03 1.422e+04 -3.471 0.000521 \*\*\*  
## PatientCat2:wordposn\_5 -4.850e-03 1.922e-03 1.422e+04 -2.524 0.011611 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) PtntC2 wrdp\_2 wrdp\_3 wrdp\_4 wrdp\_5 PC2:\_2 PC2:\_3 PC2:\_4  
## PatientCat2 -0.813   
## wordposn\_2 -0.206 0.167   
## wordposn\_3 -0.206 0.167 0.500   
## wordposn\_4 -0.206 0.167 0.500 0.500   
## wordposn\_5 -0.206 0.167 0.500 0.500 0.500   
## PtntCt2:w\_2 0.168 -0.205 -0.817 -0.408 -0.408 -0.408   
## PtntCt2:w\_3 0.168 -0.205 -0.408 -0.817 -0.408 -0.408 0.500   
## PtntCt2:w\_4 0.168 -0.205 -0.408 -0.408 -0.817 -0.408 0.500 0.500   
## PtntCt2:w\_5 0.168 -0.205 -0.408 -0.408 -0.408 -0.817 0.500 0.500 0.500

emm <- emmeans(mpos2, pairwise ~ wordpos | PatientCat, pbkrtest.limit = 14060)

## Note: D.f. calculations have been disabled because the number of observations exceeds 14060.  
## To enable adjustments, add the argument 'pbkrtest.limit = 14330' (or larger)  
## [or, globally, 'set emm\_options(pbkrtest.limit = 14330)' or larger];  
## but be warned that this may result in large computation time and memory use.

## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.  
## To enable adjustments, add the argument 'lmerTest.limit = 14330' (or larger)  
## [or, globally, 'set emm\_options(lmerTest.limit = 14330)' or larger];  
## but be warned that this may result in large computation time and memory use.

pairs(emm, adjust = "tukey")

## PatientCat = 1:  
## contrast estimate SE df z.ratio p.value  
## n\_1 - n\_2 0.001994 0.00157 Inf 1.270 0.7097  
## n\_1 - n\_3 0.008001 0.00157 Inf 5.097 <.0001  
## n\_1 - n\_4 0.007272 0.00157 Inf 4.632 <.0001  
## n\_1 - n\_5 0.009200 0.00157 Inf 5.860 <.0001  
## n\_2 - n\_3 0.006008 0.00157 Inf 3.827 0.0012  
## n\_2 - n\_4 0.005279 0.00157 Inf 3.362 0.0069  
## n\_2 - n\_5 0.007206 0.00157 Inf 4.590 <.0001  
## n\_3 - n\_4 -0.000729 0.00157 Inf -0.464 0.9905  
## n\_3 - n\_5 0.001198 0.00157 Inf 0.763 0.9411  
## n\_4 - n\_5 0.001927 0.00157 Inf 1.228 0.7353  
##   
## PatientCat = 2:  
## contrast estimate SE df z.ratio p.value  
## n\_1 - n\_2 0.001181 0.00111 Inf 1.066 0.8241  
## n\_1 - n\_3 0.007817 0.00111 Inf 7.055 <.0001  
## n\_1 - n\_4 0.013942 0.00111 Inf 12.582 <.0001  
## n\_1 - n\_5 0.014050 0.00111 Inf 12.680 <.0001  
## n\_2 - n\_3 0.006636 0.00111 Inf 5.989 <.0001  
## n\_2 - n\_4 0.012761 0.00111 Inf 11.516 <.0001  
## n\_2 - n\_5 0.012869 0.00111 Inf 11.614 <.0001  
## n\_3 - n\_4 0.006124 0.00111 Inf 5.527 <.0001  
## n\_3 - n\_5 0.006233 0.00111 Inf 5.625 <.0001  
## n\_4 - n\_5 0.000108 0.00111 Inf 0.098 1.0000  
##   
## Degrees-of-freedom method: asymptotic   
## P value adjustment: tukey method for comparing a family of 5 estimates

#####################main effects######################  
myMat <- solve(t(matrix(c(1/5,1/5,1/5,1/5,1/5,-1,1,0,0,0,-1,0,1,0,0,-1,0,0,1,0,-1,0,0,0,1), nrow =5, ncol = 5))) ## matrix for word\_type  
design\_matrix <- hypr()  
cmat(design\_matrix, add\_intercept = FALSE) <- myMat  
contrasts(df2$wordpos) <- myMat[,c(2:5)]  
  
mpos = lmer(w2v ~ PatientCat\*wordpos + (1 | ID), data = df2)   
summary(mpos)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: w2v ~ PatientCat \* wordpos + (1 | ID)  
## Data: df2  
##   
## REML criterion at convergence: -55491.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.7532 -0.6359 -0.1011 0.5915 4.5683   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.0004796 0.02190   
## Residual 0.0011744 0.03427   
## Number of obs: 14330, groups: ID, 106  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 1.710e-01 3.685e-03 1.035e+02 46.404 < 2e-16 \*\*\*  
## PatientCat2 1.294e-02 4.536e-03 1.036e+02 2.853 0.005229 \*\*   
## wordpos1 -1.994e-03 1.570e-03 1.422e+04 -1.270 0.204106   
## wordpos2 -8.001e-03 1.570e-03 1.422e+04 -5.097 3.50e-07 \*\*\*  
## wordpos3 -7.272e-03 1.570e-03 1.422e+04 -4.632 3.65e-06 \*\*\*  
## wordpos4 -9.200e-03 1.570e-03 1.422e+04 -5.860 4.73e-09 \*\*\*  
## PatientCat2:wordpos1 8.128e-04 1.922e-03 1.422e+04 0.423 0.672306   
## PatientCat2:wordpos2 1.843e-04 1.922e-03 1.422e+04 0.096 0.923606   
## PatientCat2:wordpos3 -6.669e-03 1.922e-03 1.422e+04 -3.471 0.000521 \*\*\*  
## PatientCat2:wordpos4 -4.850e-03 1.922e-03 1.422e+04 -2.524 0.011611 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) PtntC2 wrdps1 wrdps2 wrdps3 wrdps4 PtC2:1 PtC2:2 PtC2:3  
## PatientCat2 -0.812   
## wordpos1 0.000 0.000   
## wordpos2 0.000 0.000 0.500   
## wordpos3 0.000 0.000 0.500 0.500   
## wordpos4 0.000 0.000 0.500 0.500 0.500   
## PtntCt2:wr1 0.000 0.000 -0.817 -0.408 -0.408 -0.408   
## PtntCt2:wr2 0.000 0.000 -0.408 -0.817 -0.408 -0.408 0.500   
## PtntCt2:wr3 0.000 0.000 -0.408 -0.408 -0.817 -0.408 0.500 0.500   
## PtntCt2:wr4 0.000 0.000 -0.408 -0.408 -0.408 -0.817 0.500 0.500 0.500

emm2 <- emmeans(mpos, pairwise ~ wordpos | PatientCat, pbkrtest.limit = 14060)

## Note: D.f. calculations have been disabled because the number of observations exceeds 14060.  
## To enable adjustments, add the argument 'pbkrtest.limit = 14330' (or larger)  
## [or, globally, 'set emm\_options(pbkrtest.limit = 14330)' or larger];  
## but be warned that this may result in large computation time and memory use.  
## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.  
## To enable adjustments, add the argument 'lmerTest.limit = 14330' (or larger)  
## [or, globally, 'set emm\_options(lmerTest.limit = 14330)' or larger];  
## but be warned that this may result in large computation time and memory use.

pairs(emm2, adjust = "tukey")

## PatientCat = 1:  
## contrast estimate SE df z.ratio p.value  
## n\_1 - n\_2 0.001994 0.00157 Inf 1.270 0.7097  
## n\_1 - n\_3 0.008001 0.00157 Inf 5.097 <.0001  
## n\_1 - n\_4 0.007272 0.00157 Inf 4.632 <.0001  
## n\_1 - n\_5 0.009200 0.00157 Inf 5.860 <.0001  
## n\_2 - n\_3 0.006008 0.00157 Inf 3.827 0.0012  
## n\_2 - n\_4 0.005279 0.00157 Inf 3.362 0.0069  
## n\_2 - n\_5 0.007206 0.00157 Inf 4.590 <.0001  
## n\_3 - n\_4 -0.000729 0.00157 Inf -0.464 0.9905  
## n\_3 - n\_5 0.001198 0.00157 Inf 0.763 0.9411  
## n\_4 - n\_5 0.001927 0.00157 Inf 1.228 0.7353  
##   
## PatientCat = 2:  
## contrast estimate SE df z.ratio p.value  
## n\_1 - n\_2 0.001181 0.00111 Inf 1.066 0.8241  
## n\_1 - n\_3 0.007817 0.00111 Inf 7.055 <.0001  
## n\_1 - n\_4 0.013942 0.00111 Inf 12.582 <.0001  
## n\_1 - n\_5 0.014050 0.00111 Inf 12.680 <.0001  
## n\_2 - n\_3 0.006636 0.00111 Inf 5.989 <.0001  
## n\_2 - n\_4 0.012761 0.00111 Inf 11.516 <.0001  
## n\_2 - n\_5 0.012869 0.00111 Inf 11.614 <.0001  
## n\_3 - n\_4 0.006124 0.00111 Inf 5.527 <.0001  
## n\_3 - n\_5 0.006233 0.00111 Inf 5.625 <.0001  
## n\_4 - n\_5 0.000108 0.00111 Inf 0.098 1.0000  
##   
## Degrees-of-freedom method: asymptotic   
## P value adjustment: tukey method for comparing a family of 5 estimates