R Script

Student ID :35764646

library(readr)

library(dplyr)

library(lme4)

library(nlme)

library(lmerTest)

library(psych)

spanishr<- read\_csv("davies-2007-spanish-words-selected-norms\_2021-07-20.csv")

spanishr$RAN\_combined <- spanishr$RAN\_colours\_acc/spanishr$RAN\_colours\_times

spanishr$PROLEC\_word\_combined <- spanishr$PROLEC\_R\_word\_reading\_acc/spanishr$PROLEC\_R\_word\_times

spanishr$PROLEC\_nonword\_combined <- spanishr$PROLEC\_R\_nonword\_acc/spanishr$PROLEC\_R\_nonword\_times

spanishr$logfrequency <- log10(spanishr$esp.count/462.611693)

spanishr$zlogfrequency <- scale(spanishr$logfrequency, scale = TRUE, center = TRUE)

spanishr$RAN\_combined\_std <- scale(spanishr$RAN\_combined, scale = TRUE, center = TRUE)

spanishr$PROLEC\_nonword\_combined\_std <-scale(spanishr$PROLEC\_nonword\_combined, scale= TRUE, center =TRUE)

spanishr$esp.Lev\_N\_std<-scale(spanishr$esp.Lev\_N, scale=TRUE, center= TRUE)

spanishr$esp.num\_letters\_std <- scale(spanishr$esp.num\_letters , scale=TRUE , center=TRUE)

spanishr <-spanishr[!(spanishr$subj\_identifier=="adrm1166"),]

accuracy\_summary <- spanishr %>%

group\_by(subj\_identifier) %>%

summarise(accuracy\_sum = sum(accuracy))

boxplot(accuracy\_summary$accuracy\_sum, main="Scores of Acuracy", ylab="Reading Scores")

select(spanishr, accuracy, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std) %>% pairs.panels(

method = "pearson", # correlation method

hist.col = "#00AFBB",

density = TRUE # show correlation ellipses

)

spanishr.glmm0 = lmer(accuracy~(1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm0)

spanishr.glmm1 = lmer(accuracy~ RAN\_combined\_std + (1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm1)

anova(spanishr.glmm0, spanishr.glmm1)

spanishr.glmm2 = lmer(accuracy~ RAN\_combined\_std+ PROLEC\_nonword\_combined\_std + (1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm2)

anova(spanishr.glmm1, spanishr.glmm2)

spanishr.glmm3 = lmer(accuracy~ RAN\_combined\_std + zlogfrequency + PROLEC\_nonword\_combined\_std + (1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm3)

anova(spanishr.glmm2, spanishr.glmm3)

spanishr.glmm4= lmer(accuracy~ esp.Lev\_N\_std +RAN\_combined\_std + zlogfrequency + PROLEC\_nonword\_combined\_std + (1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm4)

anova(spanishr.glmm3, spanishr.glmm4)

spanishr.glmm5=lmer(accuracy~ esp.num\_letters\_std + esp.Lev\_N\_std +RAN\_combined\_std + zlogfrequency + PROLEC\_nonword\_combined\_std + (1|palabra) + (1|subj\_identifier), data = spanishr)

summary(spanishr.glmm5)

anova(spanishr.glmm4, spanishr.glmm5)

anova(spanishr.glmm0 , spanishr.glmm5)

select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std) %>% cor()%>% corrplot::corrplot()

select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std) %>% cor() %>% corrplot::corrplot()

p.mat <- corrplot::cor.mtest(select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std))$p

col <- colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))

corrplot::corrplot(cor(select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std)), method = "color", col = col(200),

type = "upper", number.cex = .5,

addCoef.col = "black", # Add coefficient of correlation

tl.col = "black", tl.srt = 90, tl.cex = .5, # Text label color and rotation

# Combine with significance

p.mat = p.mat, sig.level = 0.05, insig = "blank",

# hide correlation coefficient on the principal diagonal

diag = FALSE)

data.frame(cor(select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std)))

spanishr\_cormatrix <- data.frame(cor(select(spanishr, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std)))

se <- sqrt(diag(vcov(spanishr.glmm5)))

(tab <- cbind(Est = fixef(spanishr.glmm5), LL = fixef(spanishr.glmm5) - 1.96 \* se, UL = fixef(spanishr.glmm5) + 1.96 \* se))

exp(tab)

hist(spanishr$PROLEC\_nonword\_combined\_std)

table(spanishr$accuracy)

accuracy\_summary2 <- spanishr %>%

group\_by(subj\_identifier) %>%

mutate(accuracy\_sum = sum(accuracy)) %>%

distinct(subj\_identifier, .keep\_all=TRUE)

data.frame(cor(select(spanishr, esp.num\_letters\_std, RAN\_combined\_std, PROLEC\_nonword\_combined\_std, zlogfrequency, esp.Lev\_N\_std))) %>% round(2)