Jul.6

Xinyu Zhang

7/6/2020

Generating the Files

```
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
```

```
#reading in listening test results:
listener1 <- read.csv("Results_AK.csv", header = TRUE, sep = ";")
listener2 <- read.csv("Results_KS.csv", header = TRUE, sep = ";")

#renaming the two answers, removing the column "Num",
#and adding a dummy variable (= speaker + T) as an identifier for when joining two datasets:
SLP1 <- listener1 %>%
  dplyr::select(-Num) %>%
  rename(Roughness = Answer1, Breathiness = Answer2)
SLP1$dummy <- paste(SLP1$Speaker, "-", SLP1$T)
#nrow(SLP1) #96
SLP2 <- listener2 %>%
  rename(Roughness = Answer1, Breathiness = Answer2)%>%
  dplyr::select(-Num)
SLP2$dummy <- paste(SLP2$Speaker, "-", SLP2$T)
#nrow(SLP2) #96

AVQI <- read.csv("AVQI_results_new.csv")
#adding an dummy variable (= speaker + T) as an identifier when joining two datasets:
AVQI$dummy <- paste(AVQI$Speaker, "-", AVQI$T)
#converting "Jitter" as a numeric value from a factor (this introduces new NAs):
AVQI$Jitter <- as.numeric(levels(AVQI$Jitter))[AVQI$Jitter]
```

```
## Warning: NAs introduced by coercion
```

```
#defining the mean and sd of the Brethiness and Roughness scores:
library(dplyr)
mB1 <- mean(SLP1$Breathiness)#AK
sdB1 <- sd(SLP1$Breathiness)#AK
mB2 <- mean(SLP2$Breathiness)#KS
sdB2 <- sd(SLP2$Breathiness)#KS
mR1 <- mean(SLP1$Roughness)#AK
sdR1 <- sd(SLP1$Roughness)#AK
mR2 <- mean(SLP2$Roughness)#KS
sdR2 <- sd(SLP2$Roughness)#KS

#adding a column containing z-values for each NLP score table
zSLP1 <- SLP1 %>%
  mutate(zBreathy = (Breathiness-mB1)/sdB1, zRough = (Roughness-mR1)/sdR1) %>%
  select(-A, -Z, -T)
zSLP2 <- SLP2 %>%
  mutate(zBreathy = (Breathiness-mB2)/sdB2, zRough = (Roughness-mR2)/sdR2) %>%
  select(-A, -Z, -T)

#joining the scores from two SLPs, by matching the dummy identifier, and removing redundant rows
zSLP <- full_join(zSLP1, zSLP2, by = "dummy") %>%
  select(-Speaker.x, -Speaker.y)

#calculating the mean of z scores for each token of recording
zSLP$mzB = rowMeans(zSLP[,c('zBreathy.x', 'zBreathy.y')], na.rm=TRUE)
zSLP$mzR = rowMeans(zSLP[,c('zRough.x', 'zRough.y')], na.rm=TRUE)

#join the two SLP results tables with AVQI by matching both Speaker and T
zAVQI_ <-inner_join(AVQI,zSLP, by = "dummy")
zAVQI <- na.omit(zAVQI_)#removing rows containing NAs

#write the file for later convenience
write.csv(zAVQI, "zAVQI.csv")
#zAVQI.csv now has the AVQI data and the SLP ratings combined and matched,
#and with the variables in the right classes
```

Step-wise model comparison

```
#decorrelating breathiness from roughness:
breathyRough <- lm(mzB ~ mzR, data = zAVQI)
mzBCorr <- fitted(breathyRough, zAVQI$mzR)
zAVQI$mzBDecorr <- (zAVQI$mzB - mzBCorr)

#using step to determine the best linear model
##roughness
zRoughModel.tmp <- lm(mzR ~ CPPS + HNR + Jitter + Shimmer + ShdB + Slope + Tilt, zAVQI)
zRoughModel <- step(zRoughModel.tmp)
```

```
## Start:  AIC=-51.72
## mzR ~ CPPS + HNR + Jitter + Shimmer + ShdB + Slope + Tilt
##
```

```
##           Df Sum of Sq    RSS     AIC
## - Shimmer  1   0.00003 46.572 -53.724
## - ShdB     1   0.17259 46.744 -53.373
## - Jitter   1   0.22333 46.795 -53.270
## - Slope    1   0.25942 46.831 -53.196
## - HNR      1   0.67469 47.246 -52.358
## <none>                 46.572 -51.724
## - Tilt     1   1.51821 48.090 -50.676
## - CPPS     1   1.75065 48.322 -50.218
##
## Step:  AIC=-53.72
## mzR ~ CPPS + HNR + Jitter + ShdB + Slope + Tilt
##
##          Df Sum of Sq    RSS     AIC
## - Jitter  1   0.22375 46.795 -55.269
## - Slope   1   0.26136 46.833 -55.192
## - HNR     1   0.67780 47.249 -54.351
## <none>                 46.572 -53.724
## - ShdB    1   1.28980 47.862 -53.129
## - Tilt    1   1.54248 48.114 -52.628
## - CPPS    1   1.75157 48.323 -52.217
##
## Step:  AIC=-55.27
## mzR ~ CPPS + HNR + ShdB + Slope + Tilt
##
##         Df Sum of Sq    RSS     AIC
## - Slope  1    0.1140 46.909 -57.037
## <none>               46.795 -55.269
## - ShdB   1    1.1631 47.959 -54.936
## - Tilt   1    1.3223 48.118 -54.621
## - HNR    1    1.4114 48.207 -54.446
## - CPPS   1    1.5881 48.384 -54.098
##
## Step:  AIC=-57.04
## mzR ~ CPPS + HNR + ShdB + Tilt
##
##        Df Sum of Sq    RSS     AIC
## <none>              46.909 -57.037
## - ShdB  1    1.1331 48.043 -56.770
## - Tilt  1    1.2620 48.171 -56.515
## - CPPS  1    1.6226 48.532 -55.807
## - HNR   1    3.2663 50.176 -52.643
```

```
##breathiness
zBrDecorrModel.tmp <- lm(mzBDecorr ~ CPPS + HNR + Jitter + Shimmer + ShdB + Slope + Tilt, zAVQI)
zBrDecorrModel <- step(zBrDecorrModel.tmp)
```

```
## Start:  AIC=-107.13
## mzBDecorr ~ CPPS + HNR + Jitter + Shimmer + ShdB + Slope + Tilt
##
##           Df Sum of Sq    RSS      AIC
## - Tilt     1    0.0211 26.014 -109.049
## - Jitter   1    0.0583 26.051 -108.913
## - Slope    1    0.2668 26.259 -108.156
```

```
## <none>                 25.993 -107.126
## - Shimmer  1    0.6408 26.633 -106.812
## - ShdB     1    1.1638 27.156 -104.965
## - HNR      1    2.9104 28.903  -99.043
## - CPPS     1    7.5856 33.578  -84.800
##
## Step:  AIC=-109.05
## mzBDecorr ~ CPPS + HNR + Jitter + Shimmer + ShdB + Slope
##
##           Df Sum of Sq    RSS      AIC
## - Jitter   1    0.0933 26.107 -110.709
## - Slope    1    0.2470 26.261 -110.151
## <none>                 26.014 -109.049
## - Shimmer  1    0.6791 26.693 -108.601
## - ShdB     1    1.2238 27.237 -106.682
## - HNR      1    3.0243 29.038 -100.601
## - CPPS     1    7.9349 33.949  -85.758
##
## Step:  AIC=-110.71
## mzBDecorr ~ CPPS + HNR + Shimmer + ShdB + Slope
##
##           Df Sum of Sq    RSS      AIC
## - Slope    1    0.4602 26.567 -111.049
## <none>                 26.107 -110.709
## - Shimmer  1    0.7398 26.847 -110.054
## - ShdB     1    1.2559 27.363 -108.245
## - HNR      1    4.1818 30.289  -98.594
## - CPPS     1    8.9054 35.012  -84.826
##
## Step:  AIC=-111.05
## mzBDecorr ~ CPPS + HNR + Shimmer + ShdB
##
##           Df Sum of Sq    RSS      AIC
## <none>                 26.567 -111.049
## - Shimmer  1    0.6392 27.206 -110.790
## - ShdB     1    1.1559 27.723 -109.003
## - HNR      1    4.2517 30.819  -98.946
## - CPPS     1    9.9305 36.498  -82.880
```

```
#summarizing the two models
summary(zRoughModel)
```

```
##
## Call:
## lm(formula = mzR ~ CPPS + HNR + ShdB + Tilt, data = zAVQI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43690 -0.52251  0.09569  0.54768  1.43368
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.04900    1.09769  -3.689 0.000385 ***
## CPPS         0.08438    0.04782   1.764 0.081058 .
```

```
## HNR          0.09223    0.03684   2.503 0.014107 *
## ShdB         0.83508    0.56637   1.474 0.143852
## Tilt        -0.11590    0.07448  -1.556 0.123207
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.722 on 90 degrees of freedom
## Multiple R-squared:  0.3712, Adjusted R-squared:  0.3432
## F-statistic: 13.28 on 4 and 90 DF,  p-value: 1.519e-08
```

```
summary(zBrDecorrModel)
```

```
##
## Call:
## lm(formula = mzBDecorr ~ CPPS + HNR + Shimmer + ShdB, data = zAVQI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.56139 -0.31530 -0.01406  0.27219  1.52785
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.17100    0.72178  -0.237 0.813266
## CPPS         0.19700    0.03397   5.800 9.73e-08 ***
## HNR         -0.10644    0.02805  -3.795 0.000267 ***
## Shimmer      0.14588    0.09914   1.471 0.144645
## ShdB        -2.27560    1.14996  -1.979 0.050890 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5433 on 90 degrees of freedom
## Multiple R-squared:  0.4092, Adjusted R-squared:  0.383
## F-statistic: 15.58 on 4 and 90 DF,  p-value: 1.006e-09
```

ˆThe Above suggests that the best predictors for the roughness model are CPPS and HNR; the best predictors for the decorrelated breathiness model are CPPS, HNR, and ShdB.

Since the variables for two models overlap, the rebuilt (correleated-to-roughness) breathiness model will use exactly the same variables as the decorrelated model, only with the original (non-decorrelated) breathiness scores as the independent variable.

Generating the output

```
#running the models:
rough.model <- lm(mzR ~ CPPS + HNR, data = zAVQI)
summary(rough.model)
```

```
##
## Call:
## lm(formula = mzR ~ CPPS + HNR, data = zAVQI)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42005 -0.46659  0.01991  0.51344  1.54983
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.84331    0.28702  -6.422 5.81e-09 ***
## CPPS         0.08531    0.04249   2.008   0.0476 *
## HNR          0.06471    0.02879   2.248   0.0270 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7323 on 92 degrees of freedom
## Multiple R-squared:  0.3387, Adjusted R-squared:  0.3243
## F-statistic: 23.56 on 2 and 92 DF,  p-value: 5.487e-09
```

```
#the reconstructed breathiness model
#(using "mzB" instead of "mzBDecorr" as the independent variable here:
breathy.model <- lm(mzB ~ HNR + CPPS + ShdB, data = zAVQI)
summary(breathy.model)
```

```
##
## Call:
## lm(formula = mzB ~ HNR + CPPS + ShdB, data = zAVQI)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35561 -0.38687  0.05039  0.46346  1.18723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.14706    0.78587  -2.732  0.00756 **
## HNR         -0.04928    0.03011  -1.637  0.10518
## CPPS         0.27335    0.03702   7.384 7.09e-11 ***
## ShdB        -0.14861    0.46630  -0.319  0.75069
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5944 on 91 degrees of freedom
## Multiple R-squared:  0.5829, Adjusted R-squared:  0.5692
## F-statistic: 42.39 on 3 and 91 DF,  p-value: < 2.2e-16
```

Both CPPS and HNR are significant for Roughness (Adjusted R-squared = 0.3243). Only CPPS turned out to be significant for Breathiness (not including ShdB could result in a better p-value for HNR, this could be argued by adding in one more step to evaluate the BreathDecorr model), Adjusted R-squared = 0.5692.

Plots

```
library(scatterplot3d)
par(mfrow = c(1,2))
#roughness:
```

```
rough3d<- scatterplot3d(zAVQI$HNR, zAVQI$CPPS, zAVQI$mzR, xlab = "HNR", ylab = "CPPS", zlab = "Roughnes
rough3d$plane3d(rough.model)

#breathiness:
breathy3d<- scatterplot3d(zAVQI$HNR, zAVQI$CPPS, zAVQI$mzB, xlab = "HNR", ylab = "CPPS", zlab = "Breathi
breathy.model2 <- lm(mzB ~ HNR + CPPS, data = zAVQI)
rough3d$plane3d(breathy.model2)
```