Calculating Pathogen Detection Rate of Handle and Swab Samples

SECTION 0: Required Packages and Files

The following packages must be installed and loaded into the R script prior to execution.

```
library(scales)
library(rlang)
library(dplyr)
library(stringr)
library(tidyverse)
library(glue)
```

The datasets required to do the redcap partipant analysis can be downloaded from github (https://github.com/bbi-lab)

1. pathogen_detection_SCAN_SFS_dataset

Note that because the dataset is saved as a csv, opening it in Excel may result in loss in information. View the dataset via R or R Studio to avoid lost in information.

SECTION 1: Creating the Dataframes

Loading the Dataset for Detected Pathogens

Remember to set your working directary to the appropriate path.

One dataset is loaded and read:

```
crtldet <- read.csv("pathogen_detection_SCAN_SFS_dataset.csv")
```

crtldet contains both handle and swab samples. They are catagorized based on the Handle column; TRUE if it is a handle sample and FALSE if it is a swab sample.

Additionaly, the dataframe needs to include columns for each pathogen of interest. The columns are prefilled as FALSE; FALSE in pathogen detection of that particular sample.

```
ctrldet <- mutate(ctrldet, RSVA=FALSE, RSVB=FALSE, Flu_A_H1=FALSE,
                      Flu_A_H3=FALSE, Flu_B_pan=FALSE, hPIV1_hPIV2=FALSE, "hPIV3_hPIV4"=FALSE,
                      Flu_A_pan=FALSE, AP324NU=FALSE, hMPV=FALSE, RV_1of2=FALSE,
                      RV_2of2=FALSE,CoV_HKU1_CoV_NL63=FALSE,CoV_229E_CoV_OC43=FALSE,
                      EV_D68=FALSE,EV_pan=FALSE,HPeV=FALSE,
                      "C.pneumoniae"=FALSE,"S.pneumoniae"=FALSE,"M.pneumoniae"=FALSE,
                      AdV_1of2=FALSE,AdV_2of2=FALSE,HBoV=FALSE,
                      "Rnase P"=FALSE,"Xeno"=FALSE, "SARS-CoV-2" = "")
```

At the end of this section, there are one dataframe created and saved in the Global Environment.

1. crtldet

crtldet will be used in the next sections, towards calculating pathogen detection.

SECTION 2: Check Detection Status for Each Pathogen

Currently, the pathogen probes are all pre-filled to FALSE. TRUE will be indicated if the pathogen string is present in the detected column. The for loop will also check for detection of SARS-CoV-2 and change inconclusive results to negative.

```
# set corresponding columns to TRUE if string is present
for(row in 1:nrow(ctrldet)){
  for (col in 10:34)
    if(grepl(colnames(ctrldet[col]),ctrldet[row,5])){
      ctrldet[row,col] <- TRUE
    }
}

# check detection for SARS-CoV-2
# inconclusive results are treated as "Positive" cases
for (row in 1:nrow(ctrldet))
  ifelse(grepl("Positive",ctrldet[row,7])|grepl("Inconclusive",ctrldet[row,7]),ctrldet[row,35]
         <- "TRUE", ctrldet[row,35] <- "FALSE")
```

SECTION 3: Creating and Filling Bins

Bins needs to be created and filled for reach pathogen and/or control. Additionally, the data needs to be distinguished between handle versus swab.

```
# select columns of interest
detbin <- select(ctrldet, MBsample_id, sample_id, rack_a_nickname, oa_crt, covid_crt, Handle)

# create pathogen binnings and pre-set values to be FALSE
detbin <- mutate(detbin, "SARS-CoV-2"="", RSV=FALSE, "Flu A"=FALSE, "Flu B"=FALSE,
                 "Flu C"=FALSE, hPIV=FALSE,hMPV=FALSE, RV=FALSE, "Seasonal CoV"=FALSE,
                 EV=FALSE, HPeV=FALSE, "C.pneumoniae"=FALSE, "S.pneumoniae"=FALSE,
                 "M.pneumoniae"=FALSE, AdV=FALSE, HBoV=FALSE, "Rnase P"=FALSE,
                 "Xeno"=FALSE)

# separate dataframes into handles versus swabs
detbinH <- subset(detbin, Handle==TRUE)
detbinS <- subset(detbin, Handle==FALSE)
ctrldetH <- subset(ctrldet, Handle==TRUE)
ctrldetS <- subset(ctrldet, Handle==FALSE)
```

Once the bins are created, they can now be filled with the appropriate values.

The following is to fill the bins for handle samples:

```r
#SARS-CoV-2
for (row in 1:nrow(ctrldetH))
{detbinH[row,"SARS-CoV-2"] <- ctrldetH[row,"SARS-CoV-2"]}
#RSV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$RSVA[row] == TRUE | ctrldetH$RSVB[row] == TRUE,
         detbinH[row,"RSV"] <- TRUE,detbinH[row,"RSV"] <- FALSE)
#Flu A
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$Flu_A_H1[row] == TRUE | ctrldetH$Flu_A_H3[row] == TRUE | ctrldetH$Flu_A_pan[row] == TR
         detbinH[row,"Flu A"] <- TRUE,detbinH[row,"Flu A"] <- FALSE)
#Flu B
for (row in 1:nrow(ctrldetH))
{detbinH[row,"Flu B"] <- ctrldetH[row,"Flu_B_pan"]}
#Flu C
for (row in 1:nrow(ctrldetH))
{detbinH[row,"Flu C"] <- ctrldetH[row,"AP324NU"]}
#hPIV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$hPIV1_hPIV2[row] == TRUE | ctrldetH$hPIV3_hPIV4[row] == TRUE,
         detbinH[row,"hPIV"] <- TRUE,detbinH[row,"hPIV"]<-FALSE)
#hMPV
for (row in 1:nrow(ctrldetH))
{detbinH[row,"hMPV"] <- ctrldetH[row,"hMPV"]}
#RV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$RV_1of2[row] == TRUE | ctrldetH$RV_2of2[row] == TRUE,
         detbinH[row,"RV"] <- TRUE,detbinH[row,"RV"] <- FALSE)
#Seasonal CoV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$CoV_HKU1_CoV_NL63[row] == TRUE | ctrldetH$CoV_229E_CoV_OC43[row] == TRUE,
         detbinH[row,"Seasonal CoV"] <- TRUE,detbinH[row,"Seasonal CoV"] <- FALSE)
#EV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$EV_D68[row] == TRUE | ctrldetH$EV_pan[row] == TRUE,
         detbinH[row,"EV"] <- TRUE,detbinH[row,"EV"] <- FALSE)
#HPeV
for (row in 1:nrow(ctrldetH))
{detbinH[row,"HPeV"] <- ctrldetH[row,"HPeV"]}
#C.pneumoniae
for (row in 1:nrow(ctrldetH))
{detbinH[row,"C.pneumoniae"] <- ctrldetH[row,"C.pneumoniae"]}
#S.pneumoniae
for (row in 1:nrow(ctrldetH))
{detbinH[row,"S.pneumoniae"] <- ctrldetH[row,"S.pneumoniae"]}
#M.pneumoniae
for (row in 1:nrow(ctrldetH))
{detbinH[row,"M.pneumoniae"]<-ctrldetH[row,"M.pneumoniae"]}
#AdV
for (row in 1:nrow(ctrldetH))
  ifelse(ctrldetH$AdV_1of2[row] == TRUE | ctrldetH$AdV_2of2[row] == TRUE,
         detbinH[row,"AdV"] <- TRUE,detbinH[row,"AdV"] <- FALSE)
#HBoV
```

```
for (row in 1:nrow(ctrldetH))
{detbinH[row,"HBoV"] <- ctrldetH[row,"HBoV"]}
#Rnase P
for (row in 1:nrow(ctrldetH))
{detbinH[row,"Rnase P"] <- ctrldetH[row,"Rnase P"]}
#Xeno
for (row in 1:nrow(ctrldetH))
{detbinH[row,"Xeno"] <- ctrldetH[row,"Xeno"]}
```

The following is to fill bins for swab samples:

```
#SARS-CoV-2
for (row in 1:nrow(ctrldetS))
{detbinS[row,"SARS-CoV-2"] <- ctrldetS[row,"SARS-CoV-2"]}
#RSV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$RSVA[row] == TRUE | ctrldetS$RSVB[row] == TRUE,
         detbinS[row,"RSV"] <- TRUE,detbinS[row,"RSV"] <- FALSE)
#Flu A
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$Flu_A_H1[row] == TRUE | ctrldetS$Flu_A_H3[row] == TRUE | ctrldetS$Flu_A_pan[row] == TR
         detbinS[row,"Flu A"] <- TRUE,detbinS[row,"Flu A"] <- FALSE)
#Flu B
for (row in 1:nrow(ctrldetS))
{detbinS[row,"Flu B"] <- ctrldetS[row,"Flu_B_pan"]}
#Flu C
for (row in 1:nrow(ctrldetS))
{detbinS[row,"Flu C"] <- ctrldetS[row,"AP324NU"]}
#hPIV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$hPIV1_hPIV2[row] == TRUE | ctrldetS$hPIV3_hPIV4[row] == TRUE,
         detbinS[row,"hPIV"] <- TRUE,detbinS[row,"hPIV"] <- FALSE)
#hMPV
for (row in 1:nrow(ctrldetS))
{detbinS[row,"hMPV"] <- ctrldetS[row,"hMPV"]}
#RV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$RV_1of2[row] == TRUE | ctrldetS$RV_2of2[row] == TRUE,
         detbinS[row,"RV"] <- TRUE,detbinS[row,"RV"] <- FALSE)
#Seasonal CoV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$CoV_HKU1_CoV_NL63[row] == TRUE | ctrldetS$CoV_229E_CoV_OC43[row] == TRUE,
         detbinS[row,"Seasonal CoV"] <- TRUE,detbinS[row,"Seasonal CoV"] <- FALSE)
#EV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$EV_D68[row] == TRUE | ctrldetS$EV_pan[row] == TRUE,
         detbinS[row,"EV"] <- TRUE,detbinS[row,"EV"] <- FALSE)
#HPeV
for (row in 1:nrow(ctrldetS))
{detbinS[row,"HPeV"] <- ctrldetS[row,"HPeV"]}
#C.pneumoniae
for (row in 1:nrow(ctrldetS))
{detbinS[row,"C.pneumoniae"] <- ctrldetS[row,"C.pneumoniae"]}
```

```
#S.pneumoniae
for (row in 1:nrow(ctrldetS))
{detbinS[row,"S.pneumoniae"] <- ctrldetS[row,"S.pneumoniae"]}
#M.pneumoniae
for (row in 1:nrow(ctrldetS))
{detbinS[row,"M.pneumoniae"] <- ctrldetS[row,"M.pneumoniae"]}
#AdV
for (row in 1:nrow(ctrldetS))
  ifelse(ctrldetS$AdV_1of2[row] == TRUE | ctrldetS$AdV_2of2[row] == TRUE,
         detbinS[row,"AdV"] <- TRUE,detbinS[row,"AdV"] <- FALSE)
#HBoV
for (row in 1:nrow(ctrldetS))
{detbinS[row,"HBoV"] <- ctrldetS[row,"HBoV"]}
#Rnase P
for (row in 1:nrow(ctrldetS))
{detbinS[row,"Rnase P"] <- ctrldetS[row,"Rnase P"]}
#Xeno
for (row in 1:nrow(ctrldetS))
{detbinS[row,"Xeno"] <- ctrldetS[row,"Xeno"]}
```

The pathogen columns now reflect whether the pathogen was detected for a specific sample. The next section will generate the table displaying the number of detected/not detected pathogens.

SECTION 4: Creating Table of Pathogen Detection Rates:

A table can be generated to display the number of detected/not detected by pathogen or control.

```
# "grab" names of columns from range 7 to 24
detcol <- colnames(detbin[7:24])

# build frame of the data table
TFdetected <- data.frame("Pathogen" = detcol, "Handle_Present" = 0,
                         "Handle_Not_Present" = 0, "Swab_Present" = 0,
                         "Swab_Not_Present" = 0, stringsAsFactors = FALSE)

# fill in pathogen detection count for data table
for(row in 1:nrow(TFdetected)){
  TFdetected$Handle_Present[row] <- nrow(detbinH[detbinH[,TFdetected$Pathogen[row]]==TRUE,])
  TFdetected$Handle_Not_Present[row] <- nrow(detbinH[detbinH[,TFdetected$Pathogen[row]]==FALSE,])
  TFdetected$Swab_Present[row] <- nrow(detbinS[detbinS[,TFdetected$Pathogen[row]]==TRUE,])
  TFdetected$Swab_Not_Present[row] <- nrow(detbinS[detbinS[,TFdetected$Pathogen[row]]==FALSE,])
}


# add SARS-CoV-2 negatives
TFdetected[1,3] <- nrow(detbinH)-TFdetected[1,2]
TFdetected[1,5] <- nrow(detbinS)-TFdetected[1,4]

# add columns for percentages/rates
TFdetected <- mutate(TFdetected, Hposper=0, Sposper=0)
```

SECTION 5: Calculating Signifcance of Pathogen Detection Rate between Handle and Swab:

A t-test can be performed to determine whether pathogen detection rate for handle is significantly different compared to pathogen detection rate for swab.

A table needs to be created where the controls, Xeno and RNAse P are removed. Once those controls are removed, the rates/percentage for handle and swab can be calculated for each pathogen.

```
# controls are removed from the table
ttesttable <- TFdetected %>% filter(!Pathogen %in% c("Rnase P","Xeno"))

# fills in rates fof handle and swab for each pathogen
for (row in 1:nrow(ttesttable)){
  ttesttable[row,6] <-ttesttable[row,2]/nrow(detbinH)
  ttesttable[row,7] <-ttesttable[row,4]/nrow(detbinS)
}
```

A t-test now can be performed to determine significant difference.

```
t.test(as.numeric(ttesttable$Hposper),as.numeric(ttesttable$Sposper),paired=TRUE)
```

The p-value is **0.40**, indicating that pathogen detection of handle sample versus pathogen detection of swab samples are not significantly different.

SECTION 6: Creating Final Table:

To make the table more visually appealing, pathogen columns are renamed.

```
TFdetected[TFdetected$Pathogen=="AdV",1] <- "Adenovirus"
TFdetected[TFdetected$Pathogen=="HBoV",1] <- "Bocavirus"
TFdetected[TFdetected$Pathogen=="EV",1] <- "Enterovirus"
TFdetected[TFdetected$Pathogen=="Flu A",1] <- "Influenza A"
TFdetected[TFdetected$Pathogen=="Flu B",1] <- "Influenza B"
TFdetected[TFdetected$Pathogen=="Flu C",1] <- "Influenza C"
TFdetected[TFdetected$Pathogen=="hMPV",1] <- "Metapneumovirus"
TFdetected[TFdetected$Pathogen=="hPIV",1] <- "Parainfluenza"
TFdetected[TFdetected$Pathogen=="HPeV",1] <- "Parechovirus"
TFdetected[TFdetected$Pathogen=="RSV",1] <- "Respiratory Syncytial Virus"
TFdetected[TFdetected$Pathogen=="RV",1] <- "Rhinovirus"
TFdetected[TFdetected$Pathogen=="SARS-CoV-2",1] <- "SARS-CoV-2*"
TFdetected[TFdetected$Pathogen=="Seasonal CoV",1] <- "Seasonal Coronavirus"
TFdetected[TFdetected$Pathogen=="C.pneumoniae",1] <- "C. pneumoniae"
TFdetected[TFdetected$Pathogen=="S.pneumoniae",1] <- "S. pneumoniae"
TFdetected[TFdetected$Pathogen=="M.pneumoniae",1] <- "M. pneumoniae"
```

A percent function is created and used to calculate percentage for pathogen detection for both handle and swab witin the table, TFdetected.

```
# percent function created
percent <- function(x, digits = 1, format = "f", ...) {
  paste0(formatC(100 * x, format = format, digits = digits, ...), "%")
}

# detection rate for handle and swab converted into percentage
for (row in 1:nrow(TFdetected)){
  TFdetected[row,6] <- percent(TFdetected[row,2]/nrow(detbinH))
  TFdetected[row,7] <- percent(TFdetected[row,4]/nrow(detbinS))
}
```

With those preparations performed, the new finalized table can be created. It is called pcttable.

```
# new table is created with pathogens of interest; controls are removed
pcttable <- data.frame("Pathogen"= c("Adenovirus", "Bocavirus", "Enterovirus", "Influenza A",
                                     "Influenza B", "Influenza C", "Metapneumovirus",
                                     "Parainfluenza", "Parechovirus",
                                     "Respiratory Syncytial Virus", "Rhinovirus", "SARS-CoV-2*",
                                     "Seasonal Coronavirus", "C. pneumoniae", "M. pneumoniae",
                                     "S. pneumoniae"), "Handle_Present"="", "Swab_Present"="",
                       stringsAsFactors = FALSE)

# percentage of detection rate for handle and swab filled into the table
for (row in 1:nrow(pcttable)){
  pcttable[row,2] <- paste(paste(TFdetected[TFdetected$Pathogen==pcttable[row,1],2],
                                 TFdetected[TFdetected$Pathogen==pcttable[row,1],6],
                                 sep=" ("),"",sep=")")
  pcttable[row,3] <- paste(paste(TFdetected[TFdetected$Pathogen==pcttable[row,1],4],
                                 TFdetected[TFdetected$Pathogen==pcttable[row,1],7],
                                 sep=" ("),"",sep=")")

# save table as csv
write.csv(pcttable, 'pcttable.csv')
view(pcttable)
}
```

The finalized table should appear as:

Pathogen	Handle_Present	Swab_Present
Adenovirus	1 (1.0%)	75 (0.6%)
Bocavirus	0 (0.0%)	12 (0.1%)
Enterovirus	1 (1.0%)	29 (0.2%)
Influenza A	4 (4.1%)	387 (3.3%)
Influenza B	0 (0.0%)	248 (2.1%)
Influenza C	0 (0.0%)	5 (0.0%)
Metapneumovirus	0 (0.0%)	75 (0.6%)
Parainfluenza	1 (1.0%)	50 (0.4%)
Parechovirus	0 (0.0%)	0 (0.0%)
Respiratory Syncytial Virus	2 (2.1%)	112 (1.0%)
Rhinovirus	6 (6.2%)	585 (5.0%)
SARS-CoV-2*	1 (1.0%)	113 (1.0%)
Seasonal Coronavirus	3 (3.1%)	342 (2.9%)
C. pneumoniae	0 (0.0%)	8 (0.1%)
M. pneumoniae	0 (0.0%)	28 (0.2%)
S. pneumoniae	3 (3.1%)	267 (2.3%)