Supplementary Material - Cleaning

Megan Beaudry

10/7/2019

## R Markdown

library(knitr)  
library(kableExtra)  
library(magick)

## Linking to ImageMagick 6.9.9.39  
## Enabled features: cairo, fontconfig, freetype, lcms, pango, rsvg, webp  
## Disabled features: fftw, ghostscript, x11

library(flextable)

##   
## Attaching package: 'flextable'

## The following objects are masked from 'package:kableExtra':  
##   
## as\_image, footnote

library(officer)

Supplementary Material: Data Cleaning In the raw data, detectable but not quantifiable (DNQ) had not been assigned a value. In order to perform downstream analysis, DNQs were assigned a value of the LOD95 that was calculated in October of 2019.

#load data. path is relative to project directory.  
library(RCurl)

## Loading required package: bitops

data <- getURL("https://raw.githubusercontent.com/epid8060fall2019/MeganBeaudry-Project\_WQ/master/data/raw\_data/10-7-19%20LOD95%20Table.csv")  
LOD95data <- read.csv(text = data, header = FALSE, stringsAsFactors = FALSE)  
LOD95data <- LOD95data[-c(1), ]  
colnames(LOD95data) <- c('Source', 'Marker', 'LOD95','x', 'y', 'z', 'a', 'b', 'c', 'd')  
LOD95data = subset(LOD95data, select = -c(x,y,z,a,b,c,d))

print(LOD95data)

## Source Marker LOD95  
## 2 Salmonella invA 4.94  
## 3 Arcobacter hsp60 8.22  
## 4 Campylobacter VD16S 11.96  
## 5 Human HF183 7.12  
## 6 Human HumM2 23.46  
## 7 Ruminant Rum2Bac 9.55  
## 8 Gull LeeSG 6.41  
## 9 Canada Goose CGO1 4.28  
## 10 Muskrat MuBac 11.96  
## 11 Dog Dog3 4.78

When there was no amplification in qPCR, it was recorded as “Not Detected” in the raw data. This was converted to 1/2 of the LOD for downstream analysis.

Data from E. coli, Enterococcus, fecal coliforms, and all qPCR marker concentrations were log10 transformed prior to analysis because the Shapiro-Wilk test for normality indicated the data was non-normally distributed data.